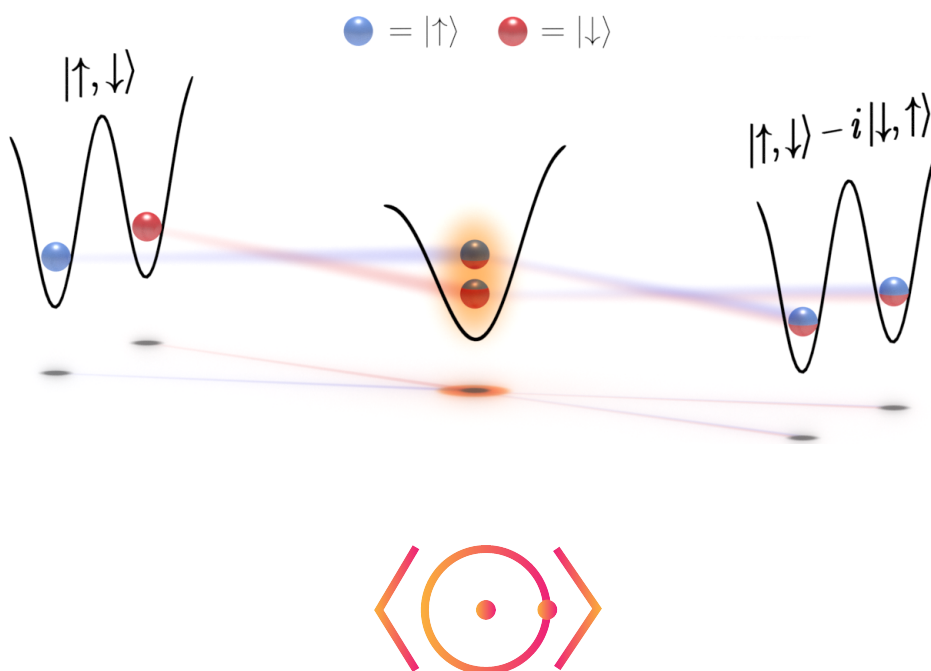


ONE, TWO, MANY — TAMING QUANTUM SYSTEMS WITH OPTIMAL CONTROL

JESPER HASSERIIS MOHR JENSEN



PhD Thesis
January 2021
Aarhus Universitet
Department of Physics and Astronomy
Supervisor: Jacob Friis Sherson

PUBLICATIONS

- [1]: J. J. Sørensen, **J. H. M. Jensen**, T. Heinzl, and J. F. Sherson. “QEngine: A C++ library for quantum optimal control of ultra-cold atoms.” In: *Computer Physics Communications* 243 (2019), pp. 135–150.
- [2]: **J. H. M. Jensen**, J. J. Sørensen, K. Mølmer, and J. F. Sherson. “Time-optimal control of collisional $\sqrt{\text{SWAP}}$ gates in ultracold atomic systems.” In: *Physical Review A* 100 (2019), p. 052314.
- [3]: **J. H. M. Jensen**, F. S. Møller, J. J. Sørensen, and J. F. Sherson. “Approximate Dynamics Lead to More Optimal Control: Efficient Calculation of Exact Derivatives.” In: *arXiv:2005.09943* (2020).
- [4]: **J. H. M. Jensen**, F. S. Møller, J. J. Sørensen, and J. F. Sherson. “Achieving fast high-fidelity control of many-body dynamics.” In: *arXiv:2008.06076* (2020).
- [5]: **J. H. M. Jensen** et al. “Crowdsourcing human common sense for quantum control.” In: *Physical Review Research* 3 (2021), p. 013057.
- [6]: **J. H. M. Jensen**, S. Z. Ahmed, C. A. Weidner et. al. “Quantum Composer: A programmable quantum visualization and simulation tool for education and research.” In: *arXiv:2006.07263* (2020).
- [7]: M. Dalgaard, F. Motzoi, **J. H. M. Jensen**, and J. F. Sherson. “Hessian-based optimization of constrained quantum control.” In: *Physical Review A* 102 (2020), p. 042612.
- [8]: C. A. Weidner, S. Z. Ahmed, **J. H. M. Jensen**, J. Sherson, and H. Lewandowski. “Investigating student use of a flexible tool for simulating and visualizing quantum mechanics.” In: *Physics Education Research Conference 2020* (2020), pp. 563–568.

Jesper Hasseriis Mohr Jensen: *One, Two, Many — Taming Quantum Systems with Optimal Control*, PhD Thesis © January 2021

Til Vibeke, Peter, Jacob & Anne
for gensidig kærlighed og støtte
og resten af min familie og venner
for alle de gode stunder

"Alt hvad der er værd at gøre, er værd at gøre godt."

ENGLISH ABSTRACT

Quantum physics is as astoundingly fascinating as a fundamental natural science as it is relevant for revolutionizing the technological landscape of society. In the last few decades especially, the character of questions asked by researchers is no longer merely if exerting precise control over tiny and fragile quantum system is *possible*, but rather *how to best do it* with respect to some desired goal.

This PhD thesis explores how a certain subset of such questions can be broadly framed and solved theoretically as optimization problems in the framework of *quantum optimal control*. The typical goal includes more than just finding perfect solutions, namely also establishing the minimal duration — as fundamentally allowed by quantum mechanics — where one such solution exists.

The herein studied physical platform is mainly ultracold atoms which is a prominent testbed for several physical processes and phenomena. Held and coherently controlled in flexible optical and/or magnetic traps, these systems also enable wide swaths of fundamentally relevant research applications. Among others, ultracold atoms exhibit scalability spanning genuine one, two, and many-body regimes as well as meanfield Bose-Einstein condensates. The body of work comprising the thesis thus covers control problems across this wide spectrum of systems. Using potent numerical instruments, this entails empirically mapping out both individual solutions and the underlying characteristic strategies in the high-fidelity, minimal duration limit.

The work presented in the thesis also covers contemporary algorithmic developments that will hopefully be useful additions to the literature. It is explored how certain implementational choices can significantly impact the feasibility of maintaining analytically exact derivatives when applying local optimization. Notably, these choices initially stem solely from independent, standard considerations in representing and solving the system dynamics and, perhaps surprisingly, the resulting working equations for the best choices do not require any new numerical machinery past the initial analytical considerations.

Lastly, the work covers alternative approaches to solving and interfacing with numerical quantum simulations and optimal control. This includes a citizen science project where control tasks are crowdsourced to nonexpert participants — they act as seeders for an optimization algorithm, accessible to them on their local device, and this human-computer interplay injects qualities that are difficult to grasp otherwise. Another type of interface, targeted towards domain (semi-)experts, is also introduced which finds use as a didactical tool in various educational and research-relevant settings.

Kvantefysikken er lige så fascinerende som fundamental naturvidenskab, som den er relevant for at revolutionere vort samfunds teknologiske landskab. Især i løbet af de seneste årtier har forskernes spørgsmål ændret karakter fra at undersøge, om hvorvidt det er blot *muligt* at kontrollere meget små og skrøbelige kvantesystemer, til nu at undersøge om *hvordan det bedst gøres* i forhold til en given målsætning.

Denne ph.d.-afhandling udforsker hvordan en delmængde af sådanne spørgsmål kan stilles og løses som optimeringsproblemer indenfor *kvanteoptimal kontrolteori*. Den typiske målsætning indbefatter som oftest mere end blot at finde en perfekt løsning, nærmere bestemt også at etablere den kortest mulige proces løbetid — hvilken er fundamentalt betinget af kvantemekanikkens love — hvor sådan en løsning findes.

Heri undersøges hovedsageligt ultrakolde atomer, og disse står til regnskab for mange fysiske processer og fænomener. Ved at holde og på kohærent vis kontrollere disse systemer med fleksible optiske og/eller magnetiske fælder, istandsættes muligheden for at udforske en bred vifte af relevante forskningsretninger. Blandt andet udviser ultrakolde atomer en særdeles skalerbarhed og forvalter derved sande et, to, og mange-legeme regi, samt middelfelts Bose-Einstein kondensater. Værkerne, der udgør denne afhandling, dækker således over kontrolproblemer, som spænder over dette brede udvalg af systemer. Ved at anvende kraftfulde numerisk instrumenter udføres empiriske undersøgelser, hvori både enkelte løsninger, og de underliggende kontrolstrategier, kortlægges. Disse gør sig samtidigt gældende indenfor det eftertragtede høj-fidelitet, minimal løbetids regi.

Arbejdet, der præsenteres her, dækker også over samtidige algoritmiske udviklinger, der forhåbentligt vil vise sig at være nyttige litteraturlitteraturskrivelser. Det undersøges bl.a., hvordan særlige implementationsvalg stærkt kan påvirke muligheden for at opretholde eksakte afledte under lokaloptimering. Disse valg opstår indledningsvis fra uafhængige, sædvanlige overvejelser, men de tilhørende, endegyldigt afledte ligninger er meget forskellige. På pudsigt vis kræves der intet nyt numerisk maskineri efter de oprindelige analytiske overvejelser.

Slutteligt dækker afhandlingen også over alternative tilgangsvinkler i forhold til at løse og *interface* med numeriske kvantesimuleringer og optimal kontrol. Dette inkluderer et borgervidenskabsprojekt, hvor kontrolopgaver *crowdsources* til lægmænd — disse sår frø for en optimeringsalgoritme, der er tilgængelig på deres lokale enhed, og dette menneske-computersammenspil tilfører kvaliteter, der ellers kan være svære at omgribe. Et andet slags *interface*, der er møntet på domæne (halv-)eksperter, introduceres også, og dette finder anvendelse som et didaktisk- og forskningsrelevant værktøj.

PREFACE

This thesis documents the majority of my scientific work and research results obtained within my four years of PhD studies from February 2017 to February 2021. Funding was provided by the European Research Council, John Templeton Foundation, Carlsberg Foundation, Lundbeck Foundation, and Innovation Fund Denmark. The work was carried out under the supervision of Jacob Friis Sherson at the Department of Physics and Astronomy, Aarhus University, Denmark.

The main focus of my studies has been on applying and developing numerical quantum optimal control techniques in solving relevant control problems in quantum physics. The balance between analytical derivations, numerical implementation, and creative exploration has been a perfect match for my personal tastes. It has been a great experience going from knowing next to nothing about the subject to at least being decent enough to write a $\mathcal{O}(10^2)$ page thesis on it.

In the early days I spent the majority of my time learning the ropes and, alongside a few others, further establishing our relatively young theoretical research group — I was only the second PhD student to join, 2-3 years after the first. This initially culminated in the release of the QENGINE which is our C++ library for numerical quantum simulation and optimal control. This library is now the canonical standard in our group.

Springing forth both from the QENGINE v.1 release and my days as a student programmer in the group (August 2015–2017), development of QUANTUM COMPOSER began. It is a visual node-based programming language interface built on top of the QENGINE. With scattered spurts of development, it is now also released in v.1. Together, the QENGINE and QUANTUM COMPOSER are software products of our *quatomic* “brand” and many of our theoretical efforts are united under this banner. The theory unit together with our colleagues in the experimental unit span the Quantum Manipulation and Measurement Group (QMMG) at the physics department.

In the later stages of my studies I have grown rather independent in my work and research projects, but also played a large part in one of our recent citizen science efforts, QUANTUM MOVES 2. There, the solution of quantum control problems are crowdsourced to nonexpert participants who are given access to an optimization algorithm implemented in the QENGINE. Together these form a so-called *hybrid intelligence*. The QUANTUM MOVES 2 project had many contributors and development of the game interface, data acquisition, etc., was done in



collaboration with the interdisciplinary [ScienceAtHome](#) unit in our group. This unit is by now quite extensive and covers many different people (I lost count of how many, but on the orders of mid tens) and topics.

Through COMPOSER and QUANTUM MOVES 2 I had the opportunity to engage in various didactical efforts through the years. Although I am by no means expert in neither citizen science, hybrid intelligence, nor didactics, I cherish the brief glimpses into these worlds. I have also been “co-supervising” students in the daily/weekly/monthly goings-on in a number of bachelors, masters, and smaller projects. This, too, has been a pleasure and it is very satisfying to teach enthusiastic faces and watch their capabilities develop.

The road to completing my studies has been a great experience both personally and professionally. I recall many obvious causes for celebration, but also, perhaps most fondly of all, seemingly minor successes and “aha!” moments where everything suddenly just clicked. That is not to say that all experiences have been equally pleasant or particularly easy. But if they were, what would be the fun in that? All in all, it is safe to say that I will always remember my time in group as nothing short of amazing. I am extremely grateful that I have been able to pursue my physics hobby to this extent.

ACKNOWLEDGMENTS

Apart from the interesting topics I have been delighted to study, I have also taken both pride and joy in working alongside so many talented and overall great people. With some it has been for but a single, fleeting moment and with others a several year stint.

First and foremost I would like to thank my supervisor prof. Jacob Sherson for giving me the opportunity, responsibility, and continual encouragement to do my PhD. I have never met someone quite as ambitious and visionary as him and it has been both impressive and inspiring to follow the group’s development over the years, and moreso taking part in this regard. I have also never met someone with such a booked calendar, but there has nevertheless nearly always been space for meeting and providing insightful feedback on manuscripts when it was necessary. My perhaps deepest appreciation lies in the friendly atmosphere, mutual respect, and trust in my being allowed to go “under the radar” for a while and resurfacing once this or that has been done. Thank you.

I would also like to extend my gratitude to prof. Klaus Mølmer both for his great character and his thoughtful contributions as a co-author on some of my publications. When I was just an enthusiast around the year 2010 my mom gifted me a book, “*Kvantemekanik: atomernes vilde verden*”, authored by Klaus. I remember reading it once late at night

and for some reason thought it becoming to memorize — despite not really understanding — the Schrödinger equation. Now I can at least justify to myself that this memory exercise was not entirely without merit.

A major collective thanks go out to all my past and present colleagues in the QMMG. In particular, I am greatly indebted to Jens Jakob Sørensen who was both a friend and initially a mentor and later a fantastic sparring partner professionally. I remember fondly our many conceptual, analytical, and technical discussions, and that time we took a week to finally get the grasp on those unwieldy Gâteaux derivatives. It was a great and fun experience working together and he was one of the determining factors in my decision to join the group as a PhD student. I also want to thank my fellow theoretical PhD student Mogens Dalgaard for several illuminating discussions as well as his company and mutual, good willed fun-poking at the office. Another major thanks go out to our post doc Carrie Weidner and PhD student Shaeema Ahmed with whom I have shared many great times, including authoring several papers. I have really appreciated your cheerful spirits and hardworking attitudes.

A collective appreciation goes out to Jens Laustsen, Robert Heck, Ottò Eliasson, Felix Motzoi, Jakub Czarkowski, Miroslav Gajdacz, Florian Korsakissok, Till Heinzl, Henrik Lund Mortensen, Poul Ejlersen, Magnus Skou, Frederik Møller, Mads Petersen, Pinja Haikka, Maja Lassen, Julie Stephansen, Janet Rafner, and many more for their direct and indirect positive impacts they have had on my time in the group.

A special thanks go out the entire ScienceAtHome team — you guys really put a unique twist on the group! Another special thanks is reserved for Fysisk Fredagsbar and its patrons for their continual supply of good times — although it has been a while, courtesy of the pandemic-that-shall-not-be-named — Leve FFB. To my family and friends inside as well as outside my academic circles: I love you.

Finally, my gratitude is extended to you, the reader, for picking up this document whether by choice or obligation. I hope you find its contents exciting or at least useful.

*Jesper Hasseriis Mohr Jensen,
Aarhus, Denmark, January 2021*

PUBLICATION SYNOPSSES AND ADHERENCE TO GSNS RULES

At the time of writing, my scientific output includes five publicly available papers [2–6] as first author, one work-in-progress paper (Ch. 5) as first (and so far only) author, one paper [1] as second author with major contributions, and two papers [7, 8] with minor contributions. The papers with minor contributions will only be mentioned in pass-

ing where applicable. The remainder will be presented and treated in accordance with the GSNS rules¹, chapter 11:

If the thesis is composed mainly of manuscripts or papers [...] the thesis must include one or several introductory sections in the student's own words (i.e., not re-using text from papers not solely written by the student) encompassing the following elements (not necessarily in this order):

1. A description of the proposed research questions in the papers
2. A summary of the results and an assessment of the applied methodologies
3. A clear description of the student's own contributions to the work, including an outline of the student's role in writing manuscripts or papers included in the thesis
4. A critical review in which the PhD student relates his or her own work to the most state-of-the-art work within the field. The PhD student must also demonstrate that he or she has an up-to-date knowledge hereof and is able to put this knowledge into a broader perspective. The review section may resemble a major review article.

These requirements are met as follows. A short synopsis of the proposed research questions (1.) and summary (2.) are listed below. The main body of the thesis then presents the content each of the manuscripts in a topically cohesive (i.e. not chronological) and interleaved order in which points (1.), (2.) and (4.) are elucidated to much wider extents. These typically span primarily over one or more chapters and go into various degrees of detail, depending on what makes sense, but all are quite comprehensive. Following the completion of primary discussions on a manuscript, a full dedicated page is inserted for reference and my contributions are explicitly detailed there (3.).

The text presented here is entirely written from "scratch" (i.e. not "copy-paste" inserted/adapted nor intentionally paraphrased). The very few occasional exceptions to this is perhaps a few paragraphs and figures in Chapters 2–3 which were also used in the progress report for the qualifying examination. This is allowed in accordance with GSNS rules. (The progress report of course also contained fledgling elements of what eventually became the individual manuscripts reported on in Chs. 6, 8, and 9, but none of those very early texts were reused.)

¹ https://phd.nat.au.dk/fileadmin/phd-nat-tech/Rules_and_regulations/NAT_Rules_and_regulations_-_September_2020_-_final_version.pdf

Nevertheless, the verbiage and style presented here may be quite similar to that in the manuscripts since I authored, depending on the particular manuscript, close to all the texts thereof and spent significant amounts of time iterating towards what I considered the best possible version in those writing processes. Additionally, relevant equations appear largely identically for obvious reasons. Figures from manuscripts are clearly indicated and most have been adapted to fit the content, structure, and style of the thesis.

All chapters present overall more material than what was originally covered in the corresponding manuscripts, typically delving into implicit and explicit “background” assumptions. These are mostly based on my own personal collection of written notes throughout the years. Additionally, certain elements only briefly mentioned in the manuscripts are expanded to extended discussion point here. All chapters contain perspectives and references to the broader context of the work as required by a subset of rule (4.). Some of the chapters presents entirely new research material, in particular Ch. 3 which is an attempt to formalize some overarching notions and Ch. 5 which includes new calculations.

Although I spent a significant amount of time of my studies on programming, numerical implementation, and technical details, these will unfortunately not be given their deserved focus in this thesis mainly due to time and space constraints. These can often be found in the manuscripts themselves. This includes, but is not limited to, the technical documentation parts of the manuscripts in Refs. [1, 6].

My personal opinion, in general, is that a certain subset of these details are not given enough emphasis or specification, likely because they are either not considered that interesting or are somewhat tedious describe. Nevertheless, I feel we should strive to close the gap between the presented results and the “whats and hows” implemented on the computer. This would also aid in the effort of bringing the “hidden arts” out in the open which, especially, would be more welcoming to newcomers of which I was one four years ago.

Along similar lines, the topical expositions are written predominantly with a relatively broad audience in mind. For example, I hope that the presentations given here will be of use to the next generation of PhD students in our group.

Below are synopses for each of the major published works and the work in progress.

Paper [1] is a introductory and technical reference for the QENGINE— our C++ library for numerical quantum simulations and optimal control. The goal of the project was to establish a high performance, well-tested, user-friendly flagship library — primarily for internal use — including several different physical models from

ultracold atomic physics and state-of-the-art optimal control algorithms. The algorithms are of the optimize-then-discretize type.

Paper [2] is the first application of the QENGINE. The goal is to effect a two-particle $\sqrt{\text{SWAP}}$ gate based on collisional interactions in an optical lattice. This was done by devising and improving optimal control results and taking into account several typically ignored aspects that gives a more realistic description of the physics.

Paper [3] explores the implications of following a discretize-then-optimize approach when calculating analytically exact control derivatives for central working equations. In taking explicitly into account the specific choice of numerical propagator and representation we find that the Suzuki-Trotter-type turns out to be clearly the best choice. Out of several standard propagators, these are the only ones not hampered by exponential scaling in ensuring exactness. Exactness is vital to any local optimization methodology.

Paper [4] deals with high-fidelity optimal control of quantum many-body systems. We drive the superfluid-Mott insulator transition of the Bose-Hubbard model in an optical lattice. The system sizes under considerations are completely beyond the reach of exact diagonalization, wherefore a matrix product state tensor network description is required. We employ an exact derivative-based control algorithm (as per Ref. [3]) to this paradigmatic problem from the many-body realm and find significantly improved quantitative and different qualitative results compared to previously.

Paper [5] reports on crowdsourcing certain quantum optimal control problems (single-particle and meanfield BEC) to citizen scientists through our game QUANTUM MOVES 2. The crowdsourcing takes place on two levels: (1) the players act as seeding mechanisms and (2) can subsequently optimize solutions on their local device using an embedded QENGINE algorithm. As a baseline test for feasibility we compare their results to uniform random seeding in several ways and find certain quantitative improvements. Future possibilities and directions for this type of citizen science are highlighted and potential gains and necessary considerations discussed. The work does not in any way suggest that humans can or should replace algorithms, but rather that a closer cooperation between these might provide certain qualities not easily grasped by a fully expert-driven algorithmic approach.

Paper [6] is an introductory reference article to QUANTUM COMPOSER—our visual programming interface for numerical quantum mechanics and optimal control. The application is a built on top of the QENGINE. This paper also describes different relevant use cases,

most prominently as a didactical tool in the educational setting but also as a research-assisting tool for domain (semi)-experts.

Chapter 5 is unpublished work which extends the analytically exact derivatives of Ref. [3] to meanfield BEC. The calculations are verified and, in comparison to the non-interacting case, these are much more involved and never exactly coincides with “usual” equations from the standard approach in the field. This implies that all meanfield BEC control to date has been employing inexact gradients. The chapter also details the steps necessary to solve state transfers between arbitrary meanfield states and describes a possible application based on a recently published control problem.

CONTENTS

1	INTRODUCTION	1
2	ULTRACOLD NEUTRAL ATOMS	7
2.1	Why Ultracold Atoms?	7
2.2	Trapping	12
2.3	Physical Modeling in 3D	17
2.4	Physical Modeling in 1D	32
2.5	System Models Considered in the Thesis	35
3	THE ELEMENTS OF QUANTUM OPTIMAL CONTROL	43
3.1	The State Transfer Problem	45
3.2	The Control Landscape	50
3.3	Traversing the Landscape by Iterative Optimization . .	56
3.4	Seeding Strategies	59
3.5	Control Parametrization	63
3.6	Cost Functionals and Handling constraints	65
3.7	Further Developments and Concepts	70
3.8	Perspectives to Other Areas of Quantum Control . . .	74
4	DERIVATIVE-BASED LOCAL OPTIMIZATION	81
4.1	Local Searching — Taking the Step	81
4.2	Imposing Discretization — A Tale of Two Cities	90
4.3	Optimize-then-discretize	93
4.4	Discretize-then-optimize	104
4.5	Comparison of Derivatives	113
4.6	Perspectives	120
	MANUSCRIPT [1] – QEngine: A C++ library for quantum optimal control of ultracold atoms	126
	MANUSCRIPT [3] – Approximate Dynamics Lead to More Optimal Control: Efficient Calculation of Exact Derivatives	127
5	BOSE-EINSTEIN CONDENSATE PREPARATION WITH EXACT GRADIENTS	129
5.1	Numerics of Meanfield BEC	129
5.2	Double Twin Beam Source Initialization	132
5.3	Exact Analytical Gradients for Meanfield BEC	139
6	TWO-PARTICLE $\sqrt{\text{SWAP}}$ COLLISION GATES	147
6.1	Introduction to the Manuscript	147
6.2	Manuscript Results	148
6.3	Perspectives	157
	MANUSCRIPT [2] – Time-optimal control of collisional $\sqrt{\text{SWAP}}$ gates in ultracold atomic systems	161
7	MANY-BODY MATRIX PRODUCT STATE CONTROL	163
7.1	The Curse of Dimensionality	163
7.2	Breaking the Curse	166
7.3	Tensor Networks	167

7.4	MPS Time Evolution and Efficient Exact Derivatives . . .	180
7.5	Introduction to the Manuscript	183
7.6	Manuscript Results	184
7.7	Perspectives	190
	MANUSCRIPT [4] – Achieving fast high-fidelity control of many-body dynamics	194
8	CROWDSOURCING QUANTUM OPTIMAL CONTROL: QUAN- TUM MOVES 2	195
8.1	Control Problem Complexity	195
8.2	Citizen science	197
8.3	Introduction to the Manuscript	199
8.4	Manuscript Results	203
8.5	Perspectives	211
	MANUSCRIPT [5] – Crowdsourcing human common sense for quantum control	216
9	QUANTUM COMPOSER	217
9.1	Current Areas of Application	218
9.2	Future Considerations	221
	MANUSCRIPT [6] – Quantum Composer: A programmable quantum visualization and simulation tool for education and research	224
10	OUTLOOK	225
10.1	Techniques	225
10.2	Control Problems	226
10.3	Interfaces	227
	BIBLIOGRAPHY	228

ACRONYMS

GP	Gross-Pitaevskii
GPE	Gross-Pitaveskii equation
BEC	Bose-Einstein condensate
GRAPE	Gradient Ascent Pulse Engineering
PGRAPE	Player Gradient Ascent Pulse Engineering (GRAPE)
SA	Stochastic Ascent
PS	Player Seeds
RS	Random Seeds
GROUP	Gradient Optimization Using Parametrization
GOAT	Gradient Optimization of Analytic Controls
DMRG	Density Matrix Renormalization Group
DBSCAN	Density-based Spatial Clustering of Applications with Noise
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CRAB	Chopped Random Basis
DCRAB	Dressed Chopped Random Basis
STIRAP	Stimulated Raman adiabatic passage
SVD	Singular Value Decomposition
MPS	Matrix Product State

INTRODUCTION

[...] we *never* experiment with just *one*
electron
or atom
or (small) molecule.

— E. Schrödinger (1952) [9]

SCIENTIFIC progress stands on the shoulder of giants. This adage certainly pertains to the realm of quantum physics, inarguably one of the greatest and most important intellectual achievements of the 20th century. Spurred on by M. Planck at the turn of the century, the consequences of his solution to the ultraviolet catastrophe lead to the paradigmatic departure from accepting classical physics as a complete description of Nature. His groundbreaking assumption was that electromagnetic black-body radiation is emitted and absorbed only in discrete “quanta”. The decades immediately following saw what may here be termed the “zeroth” quantum revolution: the foundational formulation of quantum physics [10]. Newtonian laws and determinism were superseded by mysterious notions such as wave functions, interference effects, wave-particle duality, classically forbidden tunneling, intrinsic spin, probabilities, superposition principles, uncertainty relations, entanglement, and other phenomena with no classical counterpart. This romantic period of developments was birthed by now-household figures, including the likes of N. Bohr, A. Einstein, P. Dirac, W. Heisenberg, W. Pauli, E. Schrödinger, and many more. Although the interpretation and associated philosophical implications of quantum physics have long been a topic for discussion [11], the exceptional agreement between theory and experiment have stood the test of time.

While the enduring legacy of this celebrated gallery of giants is indisputable, understanding the ground rules is — nevertheless — only half the battle. What applications would quantum find inside and outside the laboratory [12]? Only with the following steady progress of experimental capabilities and landmark breakthroughs during the mid-to-late century, quantum-based technologies matured enough to diffuse into the society at large in what is coined the *first* quantum revolution¹ [13]. Chief among them is, perhaps, semi-conductors and derivatives thereof which underpin most of modern society in the

¹ The definition of the first revolution typically also subsumes the “zeroth” revolution, but I find the presented delineation appropriate here.

form of computers, cell phones, and so on. Other notable examples include lasers [14], GPS, and MRI scanners. A common characteristic for this first wave of applications is that they do not utilize the full suite of peculiar quantum effects mentioned above. Related to this fact, although relatively understood, they rely on basic building blocks that are “given to us by Nature” [13] and these are often bulk effects (in transistors, classical bits are represented by macroscopic electron currents). What is more, we are nearing the end of Moore’s law [15] because microchip miniaturization is reaching the point where quantum effects [16], specifically quantum tunneling, are no longer negligible. All these observations suggest that there is, indeed, plenty of room at the bottom [17].

Now, one hundred and some odd years later, we are experiencing the *second* quantum revolution [13]. Especially within the last few decades, incremental theoretical and experimental advances and refinements in electronics, cooling, trapping, interrogation time, measurements, fabrication, etc., we are now not only in a position to study and manipulate very pure quantum objects, but also *engineer* these systems from the bottom up at the atomic level [18]. In other words, our role is shifting from being passive observers to active participants. Quite clearly, the emphatical comment by Schrödinger in 1952 speaks to a very different reality, to whom ours would seem an impossibly distant dream — the quote in fact continues [9],

*“[...] In thought-experiments we sometimes assume that we do;
this invariably entails ridiculous consequences.”*

These are no longer just assumptions nor thought-experiments. From macroscopic quantum clouds of condensed bosons [19–21], man-made “artificial atoms” in superconducting [22, 23] or semiconducting [24, 25] solid-state devices, teleportation of quantum information [26–29], to quantum computation [24, 30–35] and simulation [16, 36–39], these all fully embrace and certainly rely on “quantum weirdness”. As such, technologies [40] (e.g. in sensing, communication, computation, simulation) emerging from this second revolution has an enormous potential. This is reflected in the huge sums of invested capital and resources by actors in both industry, academia, and governmental bodies [41]. Major companies (notably IBM, Google, Intel, and Microsoft) and the flourishing landscape of smaller quantum startups are driving the commercialization [42, 43] of quantum products in scientific collaboration with the global academic research community whereas programs such as the European Quantum Flagship [44], the US National Quantum Initiative [45], and several Chinese projects [46] are emerging. The developing bilateral ecosystem of multiple stakeholders will likely continue to accelerate this second revolution.

Nevertheless, most of the potential surrounding the quantum “hype” is still far from realized. Major challenges and road blocks still lie ahead for these quantum technologies and are still the subject of

continual intensive research. One such category of challenges is the following. With the ability to *control* any given quantum system — typically using external electric and magnetic fields [47] — how do we *control it to do what we want*? In other words, how do we devise and design *quantum optimal control*?

Optimal control is a branch of mathematical optimization² that deals with steering a state variable subject to some dynamical equation of motion over a duration of time such that a specified objective or cost function(al) is optimized. The manipulatory access to steering the state variable is through one or more *controls* in the dynamics. These necessary ingredients form the basis for any application. Possibly the first modern, high-profile success story of optimal control was the 1960s Apollo 11 moon landing [48]: there, the spacecraft was the state variable, jet thrusters the controls, Newtons second law the dynamical equation, while a mix of time, fuel expenditure, and — of course — landing on the moon surface the optimization objective.

Adopting the optimal control framework in the quantum realm is in this sense fairly straightforward. The quantum state replaces the spacecraft, electromagnetic fields the jet thrusters, appropriate quantum dynamics the Newton equations, and the objective now encodes the yield of some physical process and typically some experimental constraints. Our job is then to engineer and identify time dependent *optimal controls* appropriate for implementing the desired transformation. This extremely compact, general problem statement has spawned a rich literature since its first applications in chemical reactions [49] (see also references in Refs. [50, 51]). Now, quantum optimal control is a broadly scoped, relatively mature field that has enjoyed theoretical and experimental success in numerous research areas [52], such as superconducting qubits [53–57], nuclear magnetic resonance systems [58–62], nitrogen vacancy centers [63–66], ions [67], cold molecules [68–71], and cold atoms [72–78], to name a few.

Several applications are indeed related to the prospective technologies expected from the second quantum revolution, but that does not preclude applications in contexts of fundamental basic research. Put simply, practically *any* stage of a given experimental sequence can be interpreted as an optimization problem when viewed through the appropriate lens. Additionally, as is in fact the norm, what may be considered fundamental science at one point may later turn into usable technology [79]. Although always principally united under the same banner — the problem formulation above — the organic evolution of optimal control within different quantum subdisciplines has lead to wide variety of research questions, philosophies, tools, approaches,

² Sometimes also called mathematical “programming”. This terminology has historic roots before the advent of computers which is what we nowadays connect with use of the word “programming”. This is a somewhat confusing — especially since we often solve mathematical programs using computer programs(!).

and techniques.

In this thesis, as suggested by the title, I present my own personal foray into this exciting domain. The most precise specification of this endeavor would be something like *numerical iterative derivative-based open-loop quantum optimal control of closed systems*. Supposedly, that just does not quite roll off the tongue as easily as simply *quantum optimal control* and the additional qualifiers will instead be elaborated later. The most important thing to clarify here is that the *numerical* really means *numerical (once many appropriate analytical considerations and calculations are accounted for)*.

In essence, the body of work presented here can therefore be understood as the union between quantum dynamics and optimization, both containing certain independent as well as interlinking analytical and numerical aspects. For example, unitary dynamics of any system can always be formally described in terms of a time-ordered exponential operator integral, but this is virtually always subject to some appropriate analytical approximation(s), depending on the system, which then turns out to have a surprising consequences for the optimization algorithmic formulation and numerical implementation. In other cases, it is the very fact that a Hilbert space scales exponentially with its constituent elements — thus severely limiting numerical treatment — that requires a completely new analytical framework for representing and calculating quantum systems beyond what is otherwise feasible.

The two primary topics of this thesis is therefore the formulation and application of quantum optimal control techniques to a suite of interesting problems. Despite being origin points for these, the technological and fundamental sciences described above will primarily serve as a backdrop and will thus be described only to a subjectively feasible and appropriate extents. A secondary, yet related topic to the primary ones is the development of novel interfaces to solve and work with numerical quantum mechanics and optimal control.

Our research group also has an experimental branch [80–85] dealing with high-resolution imaging of trapped ultracold rubidium 87 bosonic atoms in various shapes, forms, and sizes, and these types of systems therefore represent the majority of the herein studied control problems. Nevertheless, it is stressed that, e.g., the presented optimization methodologies apply equally well to any other physical platforms.

The covered works and manuscripts are distributed to provide a cohesive narrative sorted by topic rather than chronology. The intent is to illuminate common threads and trends as mentioned in the preface (brief synopses of each work is also given individually there). Each chapter described below indicates roughly the corresponding *primary* manuscript(s) — note that this is just a guiding aid as most chapters could otherwise be ascribed so many of the manuscripts that the in-

tended approximate delineation becomes meaningless. Additionally, further breakdowns of the content can be found in each chapter as appropriate.

The thesis is structured as follows.

- Chapter 2 motivates ultracold neutral atoms as an interesting platform for study, describes several interesting phenomena, and derives physical models used throughout the thesis ([1]).
- Chapter 3 gives a comprehensive overview of the elements of quantum optimal control (terminology, goals, techniques, etc.), relevant for its broad application in quantum physics.
- Chapter 4 zooms in on local optimization methodologies, discusses temporal discretization, and derives and compares analytically exact working equations ([1, 3]). A list of named standard optimal control algorithms are discussed.
- Chapter 5 continues some of the calculations from the previous chapter to also include meanfield Bose-Einstein condensate (BEC) dynamics, drastically altering the final equations. These are verified numerically. General research proposals and concrete problems for future study are given. A few preliminary results are shown.
- Chapter 6 presents considerations and results in optimally driving an two-particle entangling $\sqrt{\text{SWAP}}$ gate in the high-fidelity, minimal-duration limit ([2]).
- Chapter 7 motivates and mathematically introduces matrix product states, capable of describing genuine many-body systems, in the ansatz of which optimal control is applied to solve the paradigmatic Bose-Hubbard phase transition in the high-fidelity, minimal-duration limit ([4]).
- Chapter 8 explores the crowdsourcing of optimal control problems to nonexpert participants that, in conjunction with a local optimizer, solves certain single-particle and BEC problems in the high-fidelity, minimal-duration limit ([5]).
- Chapter 9 introduces an interface for numerical quantum simulations and optimal control which finds use in educational as well as research-assisting contexts ([6]).
- Chapter 10 provides an outlook with respect to technique development, control problems, and interfaces.

ULTRACOLD NEUTRAL ATOMS

THE main topic of this thesis is optimal control of quantum systems. That is, the engineering of time dependent control fields that implements a desired physical transformation. A natural prerequisite in this regard is to have a suite of accurate models for describing the state of affairs on the quantum level. Thus, the purpose of this chapter is to elucidate different instances of particular interest in ultracold neutral atomic systems.

The chapter is structured as follows.

Section 2.1 begins by motivating ultracold atoms as an attractive platform for quantum computation, simulation, and information as well as more fundamental studies about the nature of matter

Section 2.2 briefly introduces a few different kinds of typical trapping mechanisms available in experiments

Section 2.3 describes the general N -many-body spatial theory of ultracold atoms with contact interactions, covering both first and second quantization and, among others, the remarkable phenomenon of Bose-Einstein condensation

Section 2.4 laments on why an exact three-dimensional description is generally not feasible for systems of any appreciable N and how we can instead approximately consider a (not always *a priori* trivial) quasi-one-dimensional reduction typically achieved by separating variables

Section 2.5 summarizes for reference the variety of models relevant for the thesis and derives the Bose-Hubbard model in an optical lattice

2.1 WHY ULTRACOLD ATOMS?

Ultracold neutral atoms are of interest in both applied and fundamental physics contexts as evidenced by the expansive body of literature on the subject — see, e.g., Refs. [86–88] and references herein and hereto. A brief account surrounding the historical development and several relevant topics can be found in the introduction of Ref. [89].

Gases of bosonic and fermionic nature are now routinely cooled down through a variety different cooling techniques to, in rough numbers [90–92], temperatures typically ranging from tens or hundreds of nanokelvin to a few microkelvin. At these very low temperatures,

the quantum nature of matter becomes evident and several of these emerging effects, properties, and phenomena may be harnessed. The cooled atoms are typically loaded into optical or magnetic trapping geometries with typical length scales of hundreds of nanometers to a few micrometers¹ and typical dynamical time scales on the order of tens or hundreds of microseconds to milliseconds. Depending on several factors, such as the atomic species, state symmetry, type of trap, etc., one may realize relatively pristine systems extending from massive bulk clouds consisting of millions of atoms [93] all the way down to the single-particle level [94]. The subsequent experimental stages largely depend on the intent and designation of the experiment, ranging from, e.g., quantum computation, simulation, and information processing to more fundamental science questions about the peculiarities of Bose-Einstein condensation and degenerate Fermi gases — we shall primarily concern ourselves with bosons in this thesis with only haphazard mention of fermions. In the following subsections we briefly expand on a few not-unrelated areas of application and exemplify connections to quantum optimal control.

2.1.1 *Quantum computing, simulation, and information processing*

A classical computer operates with registers of logical bits where each bit can either be in a 0 or 1 state and never “in between”. The qubit [95] (quantum bit), on the other hand, can be in any state of the form $|q\rangle = c_0|0\rangle + c_1|1\rangle$ with $|c_0|^2 + |c_1|^2 = 1$ due to fundamental quantum superposition principles. The gains in *quantum computing* comes from the fact that a register of N qubits, $|q_1, q_2, \dots, q_N\rangle$, can store and (potentially) process exponential amounts of information compared to classical bits [16]. The typical quantum circuit paradigm is grounded in applying distinct unitary quantum gates $\mathcal{G}$ to one or two qubits in the register at a time ($\mathcal{G}_i |q_1, \dots, q_N\rangle = |q_1, \dots, \mathcal{G}_i(q_i), \dots, q_N\rangle$ and $\mathcal{G}_{i,j} |q_1, \dots, q_N\rangle = |q_1, \dots, \mathcal{G}_{i,j}(q_i, q_j), \dots, q_N\rangle$, respectively). Any desired computation can then principally be implemented by weaving together a predetermined sequence of gates, and once this *quantum algorithm* [96, 97] has been carried out the resulting quantum state can ideally be read out and interpreted classically without ambiguity. The algorithms themselves are abstract in the sense that they are principally independent of the platform, that is, the logical qubits can be encoded $|0\rangle, |1\rangle \leftrightarrow |a\rangle |b\rangle$ in any appropriate physical quantum states $|a\rangle |b\rangle$ that are sufficiently isolated and mutually addressable. Typically cited tasks in which one can expect a quantum advantage/-supremacy [98] over classical computing are, e.g., Shor’s algorithm for prime number factoring in cryptography and Grover’s algorithm in unstructured database searching [96, 97]. Their speedups are exponential and quadratic compared to classical computing, respectively, and

¹ The width of a human hair is on the order of one hundred micrometers for reference.

this correctly hints at the fact that not all algorithms equally gainful, much less at all, to implement on a quantum computer. This is clearly a non-exhaustive list and it is not the intent of this thesis to cover even these in any detail.

In delightfully concrete terms, the basic necessary ingredients for quantum computing were succinctly distilled into the five (plus two) famous criteria by DiVincenzo and colleagues almost exactly two decades ago [33]:

- (1) well-characterized, scalable qubits,
- (2) register initialization,
- (3) long coherence times (much longer than the gate operation time),
- (4) a universal set of quantum gates, and
- (5) qubit readout.

Alongside these (perhaps deceptively simple [99]) check boxes the two further criteria for quantum communication and general processing are

- (6) conversion between stationary and flying qubits, and
- (7) flying qubit transmission.

From these we can quite readily identify at least (2), (4), (6), and (7) as direct avenues for quantum optimal control and, more subtly, also (3) because it implies gate time should be minimized². In fact, we could have motivated the need for quantum optimal control solely based on these criteria.

Any contemplated physical platform [24] for realizing full scale, universal digital quantum computing must thus, as a minimum, “sufficiently” fulfill the DiVincenzo criteria to become ultimately viable. In particular, the fault-tolerant computer must remain operational by applying error-correction codes [102, 103] predicated on elementary gate errors/infidelities of roughly order 10^{-2} or less. This common quest for viability has represented major research directions within the respective platform communities since then [104, 105], each boasting individual advantages and disadvantages. We are currently in the era of Noisy Intermediate-Scale Quantum (NISQ) devices [35] containing (or are projected to contain in the near future) on the order of tens and low hundreds of qubits without substantial error correction protocols.

Among the many platform possibilities are optically trapped ultracold neutral atoms [18, 86, 94, 106–119] where the logical qubit may be stored either in external (motional) [120–123] or internal (e.g.

² Increasing the coherence time itself is also a possibility if, e.g., restricting dynamics to inside so-called decoherence-free subspaces [100, 101] by employing state-discrimination optimal control techniques discussed in Sec. 3.6.

hyperfine spin state) [33, 124–132] physical degrees of freedom. For a very recent and well-written state-of-the-art review, see Ref. [119]. Neutral atom computing architectures typically consists of optical lattices, tweezers, or combinations thereof and either qubit type is (1) strikingly scalable with (2) effectively deterministic preparatory protocols e.g. by “conventionally” driving the superfluid-Mott insulator transition [133] or more recently by configurable-array tweezer assembly [18]. Especially the internal physical qubits are cited for (3) their inherently long coherence times due to weak environmental and qubit-crosstalk couplings. Additionally, using “real” rather than “artificial” (e.g. superconducting or impurity spins) atoms ensures homogeneous qubit properties (1.) as they are fundamentally identical. Constructing (4) the set of universal gates — i.e. a set of unitaries capable of synthesizing *any* gate operation to arbitrary precision using only a finite sequence of the constituent gates [33, 95, 134–136] — typically consists in all the single-qubit gates and a suitable entangling two-qubit gate³. Single-qubit operations are implemented with e.g. site-selective lasers addressing the internal states and fidelities above 0.99 have been demonstrated in multiple experiments [119, 137, 138]. Examples of two-qubit candidates include the CNOT, CPHASE, and $\sqrt{\text{SWAP}}$ gates. The fidelities hereof are presently a few percent lower and are active research areas [139–147].

The initial conceptualization of *quantum simulation* [16, 38, 148] is broadly attributed to Feynman in the early 1980’s [30]. For these purposes, ultracold neutral atoms have also been a longstanding candidate in many respects for the reasons above. For example, loading neutral atoms into periodic light crystal structures mimics quite closely — i.e. it is an analog to — certain condensed matter and solid-state systems [89]. The upshot for these types of *analog* quantum simulators is that they are much more amenable for practical studies compared to the simulated system: they are much easier to interrogate and read out, contains no impurities (unless specifically included by design), scales of time and length are increased by several orders of magnitude. Among others this allows mapping out otherwise inaccessible, yet important, many-body phase diagrams, transitions, and scaling behaviors [149].

Next to the analog quantum simulator is the so-called *digital* quantum simulator. The main delineation [37, 119, 150, 151] between these is primarily that the former is continuously controlled with some specifically engineered Hamiltonian, $\hat{H}_{\text{analog}}(t) \approx \hat{H}_{\text{system}}(t)$, whereas the latter implements dynamics through applying a discrete (“stroboscopic”) sequence Hamiltonians and associated gates,

³ Although all universal sets are theoretically equivalent by definition, the universality condition itself does not comment on the actual gate depths necessary. For arbitrary synthesizations this depends among others on the qubit connectivity, decomposition properties, and the two-qubit gate which itself can rely on very distinct physical mechanisms.

$\hat{U}_{\text{digital}}(H_1)\hat{U}_{\text{digital}}(H_2)\cdots \approx \hat{U}_{\text{system}}$, on a qubit register that can then somehow be appropriately mapped to the simulated system. The latter is potentially more versatile but also harder to realize, and from its description it is not surprising that there are blurry-lined overlaps with quantum computing⁴. Although the particular semantics may or may not be ultimately meaningful, perhaps a conceptual difference lies in the type of tasks envisioned: quantum algorithms like Shor’s factoring (quantum computing) and the simulation of any quantum system by another quantum system (universal quantum simulation).

Similar to those of DiVincenzo, criteria encompassing both analog and digital quantum simulation have also been proposed [150]. In general, this class of tasks is widely considered somewhat simpler and more accessible in the near-term [153–155]. For example, the corresponding quantum devices need not necessarily be universal, require less overall control, and are generally less sensitive to errors since the readout typically consists in bulk expectation values (e.g. energy or magnetization in simulating condensed matter many-body systems with a neutral atom lattice system). Regardless of whether we consider the discrete digital quantum circuit gates or continuous analog simulations, the following fact remains true when assuming that the desired simulation mapping is faithful: we need to be able to correctly tailor the accessible Hamiltonian (or sequences of Hamiltonians) to induce a desired dynamics. This, also, quite clearly falls into the realm of optimal control.

2.1.2 Fundamental science

Akin to the case of quantum simulation, ultracold neutral atoms are also a prime platform for deepening our fundamental understanding of Nature in other ways. It is, for example, well-known that quantum objects can exhibit both particle- and wave-like properties. Light was famously found to be electromagnetic waves and capable of producing interference patterns centuries before establishing them as discrete bosonic particles in the photon picture. Since it turns out that massive particles are subject to the same kind of interpretation, it follows that they, too, exhibit similar interference patterns. Thickening the plot, bosonic particles may under appropriate circumstances undergo a phase transition into a so-called BEC defined by a macroscopic occupational pile up in a single quantum mode. This phenomenon occurs regardless of mass, yet its realization may, all else equal, be more readily studied in massive particles since they are much easier to “reasonably” slow down and control⁵.

⁴ As noted by Lloyd: “Quantum computers are Feynman’s universal quantum simulators” [16] and Blatt: “Such a universal quantum simulator is an instance of a special-purpose quantum computer” [37]. See also Refs. [96, 152].

⁵ Although it is indeed ultimately achievable for light [156, 157].

The onset of condensation in ultracold atoms has a nice phenomenological interpretation based on the thermal de Broglie matter wave length, $\lambda_T = \sqrt{2\pi\hbar^2/(mk_B T)}$. In a simple wave packet picture [90] it roughly describes the spatial extent of an atom at temperature T . Recalling from statistical mechanics that $\langle p^2 \rangle \sim \langle E_{\text{kinetic}} \rangle \sim T$, the expression for λ_T can also be understood as the position uncertainty pertaining to the atoms' momentum distribution (due to the canonical commutator relations $[\hat{r}_i, \hat{p}_j] = i\hbar\delta_{ij}$ and the resulting uncertainty relations). Condensation then occurs strictly below some critical temperature T_c as follows. When $T \gg T_c$, the particles behave like classical "billiard" balls, whizzing around with almost no spatial extent. As $T > T_c$ is decreased, the particles begin to behave more like genuine wave packets. When $T \sim T_c$, the matter wave lengths λ_T become comparable to the average distance between the atoms. At $T < T_c$ they begin to overlap appreciably and condense while at $T = 0$ they become effectively a single macroscopic quantum object.

BEC have a continually been subject of great research efforts [158, 159] since their experimental verification in alkali atoms in 1995 [19–21]. They are often the starting point for many types of studies, including ones from the previous section, as the condensation procedure itself is the main source for ultracold bosonic gases. This highlights the fact that what may be considered fundamental research now can well become a (part of) technology later.

The physical reality of the condensate can be modeled in different ways and its dynamics can be subject to optimal control for realizing, e.g., matter-wave interferometry [79, 160–163], production of momentum-entangled twin-atom pairs [74, 162–166], the formation of self-bound dipolar droplets [167], and so on.

2.2 TRAPPING

Precise manipulation of neutral atoms require robust trapping mechanisms. Depending on application and platform this is usually provided using optical dipole potentials, pure magnetic trapping, or combinations thereof. Examples hereof are explained in the following and one dimensional renditions are shown in Fig. 2.1 for running reference. The potential parameters will become controls, i.e. manipulatory access points, for the dynamics throughout.

2.2.1 Optical dipole potentials

Optical dipole potentials can be understood in varying degrees of "quantumness" [89, 91, 168]. In a classical oscillator picture a neutral atom subject to a (complex) electric field $\epsilon(\mathbf{r}, t)$ of driving frequency ω induces an atomic polarization $\mathbf{p} = \alpha \cdot \epsilon$. Here $\alpha(\omega)$ is the complex polarizability. By considering the real and imaginary parts of the

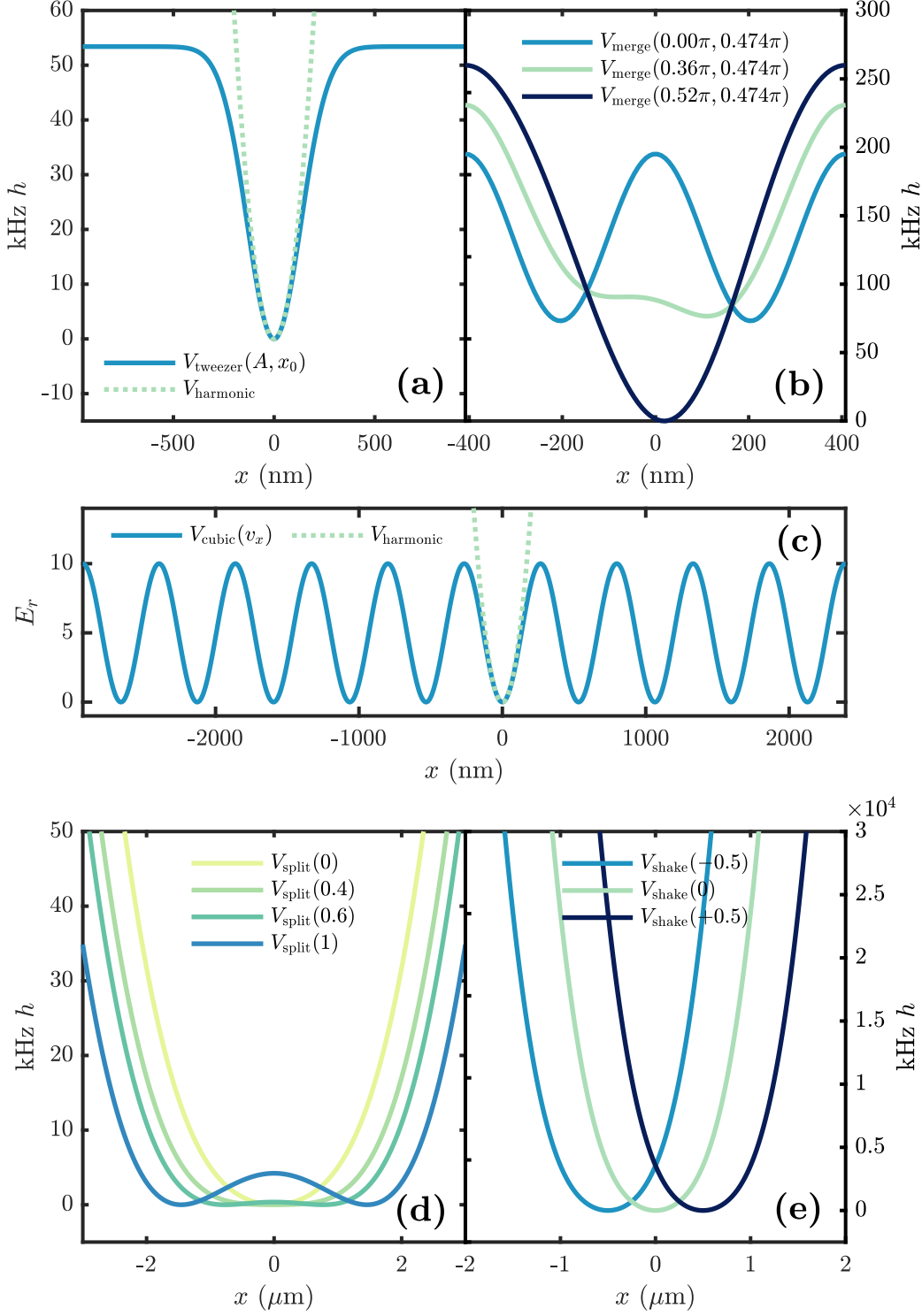


Figure 2.1: Example potentials (possibly energy shifted by a constant) in the x -direction with parameters or parameter values highlighted in legends. **(a)**: Optical tweezer Eq. (2.4) with $w_0 = \sigma = 266$ nm, $x_0 = 0$, $A = -53.42$ kHz $\cdot h$, and $\omega = 437$ kHz, $m = m_{87\text{Rb}}$ for the harmonic approximation. **(b)**: Optical lattice with site merging Eq. (2.8) with $a = 408$ nm, $V_0 = 122$ kHz $\cdot h$ and (β, θ) denoted in the legends. **(c)**: Cubic optical lattice Eq. (2.6) with $\lambda = 1064$ nm, $v_x = 10 E_r$, $E_r = \hbar^2 \pi^2 / (2a_s m) \approx 2.03$ kHz $\cdot h$ for $m = m_{87\text{Rb}}$ and $\omega = 81.2$ kHz for the harmonic approximation. **(d)**: Atom chip single-to-double well transformation Eq. (2.9) with u values denoted (see main text for value references). **(e)**: Atom chip transverse shaking Eq. (2.10) with u values denoted (see main text for value references).

interaction between the field-induced polarization and the field itself one obtains a conservative optical dipole potential U_{dip} (AC-stark shift) and a spontaneous scattering rate Γ_{sc} [91]. The polarizability α is, e.g., calculated by integrating the equation of motion of the Lorentz oscillator model [91]. These classical calculations turn out, especially for alkalis, to be good approximation to a semi-classical or even fully quantum model [91]. For the former, assuming a two-level quantum system interacting with a classical field of energy difference $\hbar\omega_0$ with negligible saturation effects and that the detuning $\Delta \equiv \omega - \omega_0$ is far off-resonance $|\Delta| \ll \omega_0$ one obtains [91]

$$U_{\text{dip}}(\mathbf{r}) = -\frac{3\pi c^2}{2\omega_0^3} \left(\frac{\Gamma}{\omega_0 - \omega} + \frac{\Gamma}{\omega_0 + \omega} \right) I(\mathbf{r}) \approx \frac{3\pi c^2}{2\omega_0^3} \frac{\Gamma}{\Delta} I(\mathbf{r}) \propto \frac{I(\mathbf{r})}{\Delta}, \quad (2.1)$$

$$\Gamma_{\text{sc}}(\mathbf{r}) = \frac{3\pi c^2}{2\hbar\omega_0^3} \left(\frac{\omega}{\omega_0} \right)^3 \left(\frac{\Gamma}{\omega_0 - \omega} + \frac{\Gamma}{\omega_0 + \omega} \right) I(\mathbf{r}) \approx \frac{3\pi c^2}{2\hbar\omega_0^3} \left(\frac{\Gamma}{\Delta} \right)^2 I(\mathbf{r}) \propto \frac{I(\mathbf{r})}{\Delta^2}, \quad (2.2)$$

where α has been eliminated, the spontaneous decay rate of the excited state is proportional to the dipole transition element, $\Gamma \sim |\langle e | \hat{\mu} | g \rangle|^2$, and we have performed the rotating wave approximation at the end. Certain modifications and additional terms may become relevant when real, multi-leveled atoms are considered but we will not concern ourselves with these details.

For large⁶ red detunings ($\Delta < 0$) the interaction is almost purely conservative and atoms are attracted to the local minima (and opposite for blue detunings $\Delta > 0$) of any inhomogeneous intensity profile. These profiles can come from a single laser (e.g. forming an optical tweezer) [169] from combining lasers (e.g. forming an optical lattices) [86, 158], very versatile digital mirror devices [82, 83, 170] as shown in Fig. 2.2, or any combination thereof.

Optical tweezer

The three-dimensional $\mathbf{r} = (x, y, z)$ potential of an optical tweezer (Gaussian beam) focused at $\mathbf{r}_0 = (x_0, y_0, z_0)$ is [91]

$$U_{\text{tweezer}}(x, y, z) = \frac{A}{\sqrt{1 + \frac{(z-z_0)^2}{z_{\text{Ray}}^2}}} \exp \left(-2 \frac{\{(x-x_0)^2 + (y-y_0)^2\}}{w^2(z-z_0)} \right), \quad (2.3)$$

where $A < 0$ is the trap depth⁷, $w(z-z_0) = w_0 \sqrt{1 + \frac{(z-z_0)^2}{z_{\text{Ray}}^2}}$, $z_{\text{Ray}} = \pi w_0 / \lambda_l$ is the Rayleigh length, λ_l is the wave length of the light, and

⁶ The weaker trapping due to large $|\Delta|$ can be compensated by increasing the intensity.

⁷ The trap depth has absorbed a number of factors. Most importantly we note $A(I) \propto I$ where $I(\mathbf{r})$ is the intensity profile. We also factored in, e.g., a factor of $1/w_0^2$ from what would otherwise be $w^2(z-z_0)$ in the denominator.

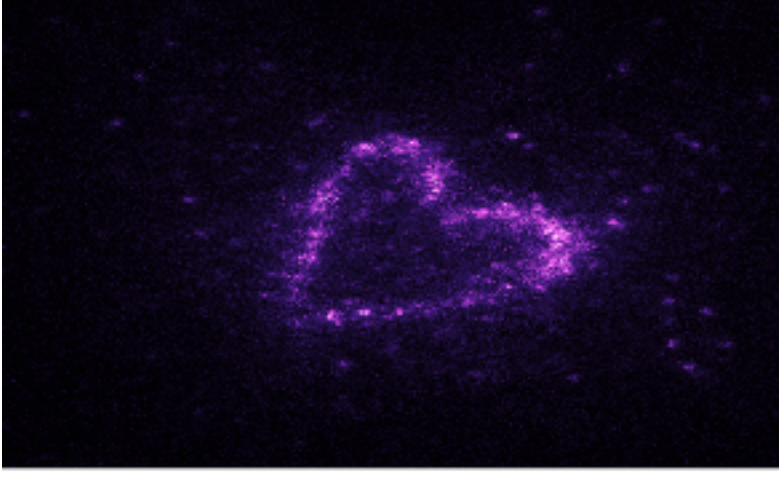


Figure 2.2: “Atomic luv”. Single-site resolved fluorescence imaging of individual atoms (color scale shows photon counts in arbitrary units) in a heart shape potential “painted” by a digital mirror device in our high-resolution imaging experiment [82–85]. The picture is included with permission from and in attribution to the experimental team.

w_0 is the beam waist which is also occasionally denoted σ . It will be prescient to record one-dimensional expressions for later

$$V_{\text{tweezer}}(x) = U_{\text{tweezer}}(x, 0, 0) = A \exp\left(-\frac{2(x - x_0)^2}{w_0^2}\right), \quad (2.4)$$

where $z_0 = y_0 = 0$ for simplicity. The harmonic approximation near the center, $V_{\text{harmonic}} = \frac{1}{2}m\omega^2(x - x_0)^2$, has a frequency of $\omega = \sqrt{-4A/(mw_0^2)}$ obtained by a second order Taylor expansion of V_{tweezer} and comparing terms. These and the following one dimensional potentials are shown in Fig. 2.1.

Optical lattices

Optical lattices are produced by interferences between superimposed lasers and this allows a wide variety of patterns [91]. The simplest three-dimensional $\mathbf{r} = (x, y, z)$ instance is the cubic lattice potential,

$$U_{\text{cubic}}(\mathbf{r}) = \sum_{q=x,y,z} v_q \sin^2 k_l q = V_x + V_y + V_z, \quad (2.5)$$

with the independent one-dimensional lattices V_q , e.g.,

$$V_{\text{cubic}}(x) = U_{\text{cubic}}(x, 0, 0) = V_x(x) = v_x \sin^2 k_l x. \quad (2.6)$$

Here, the wave number is $k_l = 2\pi/\lambda_l = \pi/a$, the laser wavelength $\lambda_l = 2a$, and the lattice separation distance a with minima at $q = 0, \pm a, \pm 2a, \dots$, and so on. In general all these quantities vary with the direction $q = x, y, z$ but for the simple cubic lattice only the trapping depth/strengths do. The trapping strengths are typically written $v_q =$

$s \cdot E_r$ in units of the recoil energy $E_r = \frac{\hbar^2 k_l^2}{2m} = \frac{\hbar^2 (\pi/a)^2}{2m} = \frac{\hbar^2 \pi^2}{2ma^2}$. The harmonic approximation near the periodic minima (site centers) has a frequency of $\omega = \sqrt{2v_x k_l^2 / m}$.

As a brief example of a more complicated interference pattern is the lattice given by [2, 140, 171]

$$U_{\text{merge}}(\mathbf{r}) = -V_z \cos^2 kz - V_0 \left[\cos^2 \left(\frac{\beta}{2} \right) \left\{ \cos^2(k_l y) + \cos^2(k_l x - \frac{\pi}{2}) \right\} + \sin^2 \left(\frac{\beta}{2} \right) \left\{ \cos(k_l y) + \cos(kx - \theta - \frac{\pi}{2}) \right\}^2 \right], \quad (2.7)$$

which in the one-dimensional x axis reads

$$V_{\text{merge}}(x) = U_{\text{lat}}(x, 0, 0) = -V_0 \left[\cos^2 \left(\frac{\beta}{2} \right) \left\{ 1 + \cos^2(k_l x - \frac{\pi}{2}) \right\} + \sin^2 \left(\frac{\beta}{2} \right) \left\{ 1 + \cos(k_l x - \theta - \frac{\pi}{2}) \right\}^2 \right]. \quad (2.8)$$

Here, V_0, V_z are trapping depths, $k_l = 2\pi/\lambda_l = \pi/a$ as before, and the parameters θ, β are adjustable phases that non-trivially transforms from a the xy -lattice plane with site separation a in the unmerged configuration into one with $a/2$ by effectively merging pairs of adjacent wells.

Other examples include superposition of two lattices with different optical frequencies, yielding a so-called superlattice [92, 131].

2.2.2 Magnetic trapping

Pure magnetic trapping in a weak $\mathbf{B}$ -field exploits the Zeeman Hamiltonian $\hat{H} = g_F \mu_B \mathbf{B} \cdot \hat{\mathbf{F}}$ where g_F is the Landé g-factor, μ_B is the Bohr magneton, and $\hat{\mathbf{F}}$ is the collective angular spin operator [172–175]. Depending on the sign of g_F states can be characterized as weak- or strong field seekers according to their spin, as they will either be attracted to local minima or maxima of the $\mathbf{B}$ field (similarly to red- and blue detuning in optical traps). See Refs. [173, 175] for topical reviews.

A useful geometry for several BEC experiments is produced by small integrated wire circuits on so-called atom-chips [172, 176]. Here, due to small wire currents, a static $\mathbf{B}$ -field is dressed by an external radio-frequency $\mathbf{B}$ -field. This geometry allows one dimensional transformations between single-well and double-well potential configurations [5, 163, 172, 177]

$$V_{\text{split}}(x) = U_{\text{split}}(x, 0, 0) = p \cdot \sqrt{(B_S(x) - B_\omega)^2 + \left(\frac{(0.5 + 0.3u)}{2B_S(x)} B_I \right)^2}, \quad (2.9)$$

where p is a constant and $B_S(x) = \sqrt{(\text{Gr} \cdot x)^2 + B_I^2}$ as well as transverse shaking

$$V_{\text{shake}}(x) = U_{\text{shake}}(x, 0, 0) = \sum_{r=2,4,6} p_r \cdot (x - u)^r, \quad (2.10)$$

valid near $|x - u| \lesssim 2 \mu\text{m}$ where p_r are fit parameters. For explicit values and units, see Refs. [1, 5] and references therein.

2.3 PHYSICAL MODELING IN 3D

We are interested in describing an N -body quantum system of indistinguishable particles in d spatial dimensions ($\mathbf{r} \in \mathbb{R}^d$). It is a widely accepted that $d = 3$ is the correct description of Nature but we shall yet track this as a variable in anticipation of later developments. For simplicity we will assume that these particles are spinless in the following and the extension is otherwise fairly straightforward.

The monolithic quantum state, $|\Psi\rangle$, encompassing all Nd degrees of freedom is a very complex and opaque object. Fortunately, we typically find recourse in its description by means of much more “intuitive” and tangible time-independent orthonormal sets of single-particle states $\{|\lambda_i\rangle\}$. These are typically eigenstates of some single-particle operator, $\hat{O}|\lambda_i\rangle = \lambda_i|\lambda_i\rangle$, and may be discretely or continuously distributed and satisfy the immensely useful identity resolutions/completeness relations

$$\text{Discrete: } 1 = \sum_i |\lambda_i\rangle \langle \lambda_i|, \quad \text{Continuous: } 1 = \int d^d \lambda |\lambda\rangle \langle \lambda|, \quad (2.11)$$

where subscripts in the continuous case are occasionally omitted from hereon and integration is over the full range of values in all d dimensions. For example, real space coordinates $\hat{\mathbf{r}}|\mathbf{r}\rangle = \mathbf{r}|\mathbf{r}\rangle$ with $\mathbf{r} \in \mathbb{R}^d$ and energy eigenstates $\hat{H}|E_i\rangle = E_i|E_i\rangle$ with $i = 0, 1, \dots$ are two commonly encountered bases on which we may seek to expand and resolve $|\Psi\rangle$.

2.3.1 State Representation — First quantization

From empirical observations it is a well-known, yet curiously remarkable, fact that all particles in Nature can be classified by the behavior of their complex valued real space wave function, $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \equiv \langle \mathbf{r}_1, \dots, \mathbf{r}_N | \Psi \rangle$, upon coordinate exchange of any two particles,

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots, \mathbf{r}_N) = \pm \Psi(\mathbf{r}_1, \dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N). \quad (2.12)$$

Occasionally we shall denote the overall state symmetry by a superscript $\Psi^\pm$. We call particles for which the $+$ ($-$) identity holds bosons

(fermions) and these symmetry requirements have immense and far reaching consequences for the development of physical theories. For example, if $\hat{P}_{ij}$ is the particle exchange operator, the only way for fermions to achieve $\hat{P}_{ij}\Psi^- = -\Psi^-$ when $r_j = r_i$ is $\Psi^- = 0$ — this is the famous Pauli exclusion principle stating that no two indistinguishable fermions can have the exact same set of quantum numbers. This is evidently not the case for bosons and this leads to, among others and under appropriate circumstances, the macroscopic quantum phenomena of Bose-Einstein condensation [158, 178] described in more detail later.

Apart from the fundamental ramifications of Eq. (2.12), we are also faced with more practical ones. Namely, even outside of symmetry considerations, Ψ quickly becomes a numerically intractable object for computations due to its Nd coordinates with normalization conditions like, e.g., $\int |\Psi|^2 d^d r_{1:N} = 1$ (here $r_{1:N}$ is a short-hand for N d -dimensional integrals). Adding the symmetry requirement makes the situation even worse as we must now also solemnly promise to properly symmetrize $|\Psi\rangle$ when expanding on the single-particle states $\{|\lambda_i\rangle\}$. That is, we must generally ensure that any permutation of the particle coordinates is consistent with the appropriate symmetry. As canonical examples, physically valid states for are on the form

$$\begin{aligned} N = 2 : \quad & (\sqrt{2})\Psi^\pm(\mathbf{r}_1, \mathbf{r}_2) = \lambda_i(\mathbf{r}_1)\lambda_j(\mathbf{r}_2) \pm \lambda_i(\mathbf{r}_2)\lambda_j(\mathbf{r}_1), \quad (2.13) \\ N = 3 : \quad & (\sqrt{6})\Psi^\pm(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \\ & \lambda_i(\mathbf{r}_1)\lambda_j(\mathbf{r}_2)\lambda_k(\mathbf{r}_3) + \lambda_i(\mathbf{r}_2)\lambda_j(\mathbf{r}_3)\lambda_k(\mathbf{r}_1) + \lambda_i(\mathbf{r}_3)\lambda_j(\mathbf{r}_1)\lambda_k(\mathbf{r}_2) \\ & \pm \lambda_i(\mathbf{r}_2)\lambda_j(\mathbf{r}_1)\lambda_k(\mathbf{r}_3) \pm \lambda_i(\mathbf{r}_1)\lambda_j(\mathbf{r}_3)\lambda_k(\mathbf{r}_2) \pm \lambda_i(\mathbf{r}_3)\lambda_j(\mathbf{r}_2)\lambda_k(\mathbf{r}_1), \\ N = 4 : \quad & \dots \\ & \vdots \end{aligned}$$

It is quite clear that the combinatoric behavior [179] with N yields increasingly tedious system descriptions induced solely by the symmetry bookkeeping — there is no “new” information and there is a seemingly large, yet strictly necessary, degree of redundancy in this formulation.

2.3.2 State Representation — Second quantization

What turns out to be a much more convenient and effective exact formalism when N increases is that of *second quantization* [179, 180]. As opposed to the *first quantization* described above, second quantization incorporates the symmetry requirements by construction and provides a much more elegant mathematical framework in which arbitrary N can be treated on the same footing.

The first basic idea is to define *occupation number* states $|n_{\lambda_i}\rangle$ in which we simply count the number (n) of atoms is in each single-particle state/mode ($|\lambda_i\rangle$). The set of all states $\prod_i |n_{\lambda_i}\rangle = |n_{\lambda_1}, n_{\lambda_2}, \dots\rangle$

consistent with some fixed $N = \sum_i n_{\lambda_i}$ forms an orthonormal basis for the so-called Fock space $\mathcal{F}^N$ on which we may expand states,

$$|\Psi\rangle = \sum_{\substack{n_{\lambda_1}, n_{\lambda_2}, \dots \\ N = \sum_i n_{\lambda_i}}} c_{n_{\lambda_1}, n_{\lambda_2}, \dots} |n_{\lambda_1}, n_{\lambda_2}, \dots\rangle. \quad (2.14)$$

Oftentimes, the explicit labeling of λ is omitted for readability, but it is important to remember that the occupation numbers are in reference to a particular set of states and in certain computations they are best written out in full. What is more, we can eliminate the fixed-particle condition by considering joining the Fock spaces for all N ,

$$\mathcal{F} = \bigoplus_{N=0}^{\infty} \mathcal{F}^N, \quad (2.15)$$

which includes the vacuous space $N = 0$ with the state $|0\rangle = |0, 0, \dots, 0\rangle = |\text{vac}\rangle$.

The second basic ingredient is the introduction of creation and annihilation operators for the single-particle modes,

$$\hat{a}_{\lambda_i} |n_{\lambda_0}, \dots, n_{\lambda_i}, \dots\rangle = \sqrt{n_{\lambda_i}} |n_{\lambda_0}, \dots, n_{\lambda_i} - 1, \dots\rangle, \quad (2.16a)$$

$$\hat{a}_{\lambda_i}^\dagger |n_{\lambda_0}, \dots, n_{\lambda_i}, \dots\rangle = \sqrt{n_{\lambda_i} + 1} |n_{\lambda_0}, \dots, n_{\lambda_i} + 1, \dots\rangle, \quad (2.16b)$$

and the definition of canonical (anti)-commutation relations for the two different particle classes,

$$\text{bosons : } [\hat{a}_{\lambda_i}, \hat{a}_{\lambda_j}^\dagger] = \delta_{i,j}, \quad [\hat{a}_{\lambda_i}, \hat{a}_{\lambda_j}] = [\hat{a}_{\lambda_i}^\dagger, \hat{a}_{\lambda_j}^\dagger] = 0, \quad (2.16c)$$

$$\text{fermions : } \{\hat{a}_{\lambda_i}, \hat{a}_{\lambda_j}^\dagger\} = \delta_{i,j}, \quad \{\hat{a}_{\lambda_i}, \hat{a}_{\lambda_j}\} = \{\hat{a}_{\lambda_i}^\dagger, \hat{a}_{\lambda_j}^\dagger\} = 0. \quad (2.16d)$$

Here $\delta_{i,j}$ is understood as a generalized Kronecker delta that is a Dirac delta function when λ_i are continuous. It is these canonical relations that ensures the symmetry requirements and thus e.g. fermionic exclusion. Note how they somewhat mimic basis orthonormality properties — it is maybe then not surprising that any operator can be described in terms of these fundamental operators [179, 180]. An important example is the counting operator, $\hat{n}_{\lambda_i} = \hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_i}$, since the occupation states are eigenstates thereof $\hat{n}_{\lambda_i} |n_{\lambda_i}\rangle = n_{\lambda_i} |n_{\lambda_i}\rangle$ (the counting operator is a special case of the density operator $\hat{\rho}_{\lambda_i, \lambda_j} = \hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_j}$). A natural extension is the total counting operator $\hat{N} = \sum_i \hat{n}_{\lambda_i}$.

The second quantization framework follows entirely from this algebra and is equivalent to first quantization. To convince ourselves that this is true we verify some of the necessary properties in the following.

Particle statistics

With the creation and annihilation operator identities we can produce any desirable quantum state starting from the vacuum as

$$|n_{\lambda_0}, n_{\lambda_1}, \dots\rangle = \prod_i \frac{(\hat{a}_{\lambda_i}^\dagger)^{n_{\lambda_i}}}{\sqrt{n_{\lambda_i}!}} |0\rangle, \quad (2.17)$$

$$\langle n_{\lambda_0}, n_{\lambda_1}, \dots| = \langle 0| \prod_i \frac{(\hat{a}_{\lambda_i})^{n_{\lambda_i}}}{\sqrt{n_{\lambda_i}!}}. \quad (2.18)$$

A caveat here is that for e.g. $\lambda = r$ the total normalization is simply $1/\sqrt{N}$, the total number of particles. Assuming $j \neq i$, we have e.g. $|\lambda_i\rangle = |1_{\lambda_i}\rangle = \hat{a}_{\lambda_i}^\dagger |0\rangle$ and $|1_{\lambda_i}, 1_{\lambda_j}\rangle = \hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_j}^\dagger |0\rangle$ for both bosons and fermions. But what happens when a mode is occupied by more than one particle? Let us count!

$$\begin{aligned} \hat{n}_{\lambda_i} |2_{\lambda_i}\rangle &= \hat{n}_{\lambda_i} \frac{1}{\sqrt{2}} (\hat{a}_{\lambda_i}^\dagger)^2 |0\rangle = \frac{1}{\sqrt{2}} \hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_i} \hat{a}_{\lambda_i}^\dagger |1_{\lambda_i}\rangle \\ &= \frac{1}{\sqrt{2}} \hat{a}_{\lambda_i}^\dagger (1 \pm \hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_i}) |1_{\lambda_i}\rangle = \frac{1}{\sqrt{2}} \hat{a}_{\lambda_i}^\dagger (1 \pm \hat{n}_{\lambda_i}) |1_{\lambda_i}\rangle \\ &= \frac{1}{\sqrt{2}} \hat{a}_{\lambda_i}^\dagger (1 \pm 1) |1_{\lambda_i}\rangle = (1 \pm 1) |2_{\lambda_i}\rangle. \end{aligned} \quad (2.19)$$

Here we used the the fundamental operator properties and canonical (anti-)commutator relations of Eq. (2.16). Therefore we must conclude that fermions correctly terminate $(\hat{a}_{\lambda_i}^\dagger)^{n_{\lambda_i}} |0\rangle = 0$ for any $n_{\lambda_i} > 1$ whereas bosons can have indefinitely many particles in the same mode. That is, effectively

$$n_{\lambda_i} = \begin{cases} 0, 1 & \text{for fermions} \\ 0, 1, 2, \dots & \text{for bosons.} \end{cases} \quad (2.20)$$

From hereon we shall be considering primarily bosons.

State normalization

Let us also verify the normalization for some of these bosonic states (omitting the explicit λ label momentarily). To this end it is convenient, as in the small calculation above, to manipulate the expressions into *normal form* where all creation operators stand to the left and all annihilation operators stand to the right — for these rearrangements, notice that

$$[\hat{a}_i, \hat{a}_j^\dagger] = \hat{a}_i \hat{a}_j^\dagger - \hat{a}_j^\dagger \hat{a}_i = \delta_{i,j} \Rightarrow \hat{a}_i \hat{a}_j^\dagger = \delta_{i,j} + \hat{a}_j^\dagger \hat{a}_i, \quad (2.21)$$

Further, let us denote the swapping of two (possibly) non-commuting operators by $\hat{a}_j \hat{a}_k^\dagger \hat{a}_i^\dagger = \left(\delta_{i,j} + \hat{a}_j^\dagger \hat{a}_i \right) \hat{a}_k^\dagger$ with the vertical lines indicating the swap-pair and use the notation $0 = \hat{\#}_i |0\rangle = \langle 0| \hat{\#}_i^\dagger$ which is useful

when there are many terms. Based on these simple rules we can then calculate

$$\langle 1_k | 1_i \rangle = \langle 0 | \hat{a}_k \hat{a}_i^\dagger | 0 \rangle = \langle 0 | (\delta_{i,k} + \hat{a}_i^\dagger \hat{a}_k) | 0 \rangle = \delta_{i,k} \langle 0 | 0 \rangle = \delta_{i,k}, \quad (2.22)$$

$$\begin{aligned} \langle 1_l, 1_k | 1_i, 1_j \rangle &= \langle 0 | \hat{a}_l \hat{a}_k \hat{a}_i^\dagger \hat{a}_j^\dagger | 0 \rangle = \langle 0 | \hat{a}_l (\delta_{i,k} + \hat{a}_i^\dagger \hat{a}_k) \hat{a}_j^\dagger | 0 \rangle \\ &= \langle 0 | \delta_{i,k} \hat{a}_l \hat{a}_j^\dagger + \hat{a}_l \hat{a}_i^\dagger \hat{a}_k \hat{a}_j^\dagger | 0 \rangle \\ &= \langle 0 | \delta_{i,k} (\delta_{j,l} + \hat{a}_j^\dagger \hat{a}_l) + \hat{a}_l \hat{a}_i^\dagger (\delta_{k,j} + \hat{a}_j^\dagger \hat{a}_k) | 0 \rangle \\ &= \langle 0 | \delta_{i,k} \delta_{j,l} + (\delta_{i,l} + \hat{a}_i^\dagger \hat{a}_l) \delta_{k,j} | 0 \rangle \\ &= \delta_{i,k} \delta_{j,l} \langle 0 | 0 \rangle + \delta_{i,l} \delta_{k,j} \langle 0 | 0 \rangle = \delta_{i,k} \delta_{j,l} + \delta_{i,l} \delta_{k,j}, \end{aligned} \quad (2.23)$$

which is 1 if $i = k, j = l$ or $i = l, k = j$ and 0 otherwise as expected and similarly

$$\langle 2_i | 2_j \rangle = \left\langle 0 \left| \frac{\hat{a}_i \hat{a}_i^\dagger}{\sqrt{2}} \frac{\hat{a}_j \hat{a}_j^\dagger}{\sqrt{2}} \right| 0 \right\rangle = \dots = \frac{1}{2} \langle 0 | \delta_{i,j} \delta_{i,j} + \delta_{i,j} \delta_{i,j} | 0 \rangle = \delta_{i,j}. \quad (2.24)$$

The second quantized states are thus properly normalized.

State symmetry

Let us establish the connection to first quantization and verify that the states are properly symmetrized. We do this by considering first some arbitrary single-particle quantum state $|\psi\rangle = \hat{a}_\psi^\dagger |0\rangle$ (i.e. not necessarily $|\lambda_i\rangle$). Then, utilizing the completeness relations Eq. (2.11), we obtain different forms for the operators by comparing the first and last equalities in the following:

$$\begin{aligned} \hat{a}_\psi^\dagger |0\rangle &= |\psi\rangle = \sum_i |\lambda_i\rangle \langle \lambda_i | \psi \rangle = \sum_i \langle \lambda_i | \psi \rangle |\lambda_i\rangle = \left(\sum_i \psi_{\lambda_i} \hat{a}_{\lambda_i}^\dagger \right) |0\rangle, \\ \hat{a}_\psi^\dagger |0\rangle &= |\psi\rangle = \int d^d \mathbf{r} |\mathbf{r}\rangle \langle \mathbf{r} | \psi \rangle = \int \langle \mathbf{r} | \psi \rangle |\mathbf{r}\rangle d^d \mathbf{r} = \left(\int \psi(\mathbf{r}) \hat{a}_\mathbf{r}^\dagger d^d \mathbf{r} \right) |0\rangle, \\ \Rightarrow \hat{a}_\psi^\dagger &= \sum_i \psi_{\lambda_i} \hat{a}_{\lambda_i}^\dagger = \int \psi(\mathbf{r}) \hat{a}_\mathbf{r}^\dagger d^d \mathbf{r}, \end{aligned} \quad (2.25)$$

$$\hat{a}_\psi = \sum_i \psi_{\lambda_i}^* \hat{a}_{\lambda_i} = \int \psi^*(\mathbf{r}) \hat{a}_\mathbf{r} d^d \mathbf{r}. \quad (2.26)$$

Here $\psi(\mathbf{r}) = \langle \mathbf{r} | \psi \rangle$ and $\psi_{\lambda_i} = \langle \lambda_i | \psi \rangle$ are the usual wave functions and state expansion coefficients, respectively, from first quantization. These rigorously derived expressions evidently correspond to just simply replacing the usual basis states by their associated canonical operators in the associated first quantization expansions. Although not really a requirement, we can interpret quite nicely the wave-particle duality of quantum mechanics — that is, quantum objects may exhibit both

wave and particle characteristics — by e.g. Eq. (2.25) since it includes both wave functions and particle operators.

As an aside for later, use we may also record some useful and common transformations between discrete $\{|\lambda_i\rangle\}, \{|\lambda'_i\rangle\}$ and continuous $\{|\mathbf{r}\rangle\}, \{|\mathbf{p}\rangle\}$ bases stemming from similar calculations,

$$\mathbf{d} \rightarrow \mathbf{d}, \quad \hat{a}_{\lambda_j}^\dagger = \sum_i \langle \lambda'_i | \lambda_j \rangle \hat{a}_{\lambda'_i}^\dagger, \quad \hat{a}_{\lambda_j} = \sum_i \langle \lambda_j | \lambda'_i \rangle \hat{a}_{\lambda'_i} \quad (2.27a)$$

$$\mathbf{d} \rightarrow \mathbf{c}, \quad \hat{a}_{\lambda_j}^\dagger = \int \lambda_j(\mathbf{r}) \hat{a}_{\mathbf{r}}^\dagger d^d \mathbf{r}, \quad \hat{a}_{\lambda_j} = \int \lambda_j^*(\mathbf{r}) \hat{a}_{\mathbf{r}} d^d \mathbf{r}, \quad (2.27b)$$

$$\mathbf{c} \rightarrow \mathbf{d}, \quad \hat{a}_{\mathbf{r}}^\dagger = \sum_i \hat{a}_{\lambda_i}^\dagger \lambda_i^*(\mathbf{r}), \quad \hat{a}_{\mathbf{r}} = \sum_i \hat{a}_{\lambda_i} \lambda_i(\mathbf{r}), \quad (2.27c)$$

$$\mathbf{d} \rightarrow \mathbf{d}, \quad \hat{a}_{\mathbf{r}}^\dagger = \int \langle \mathbf{p} | \mathbf{r} \rangle \hat{a}_{\mathbf{p}}^\dagger d^d \mathbf{p}, \quad \hat{a}_{\mathbf{r}} = \int \langle \mathbf{r} | \mathbf{p} \rangle \hat{a}_{\mathbf{p}} d^d \mathbf{p}, \quad (2.27d)$$

where $\langle \mathbf{p} | \mathbf{r} \rangle = (2\pi\hbar)^{-d/2} \exp(-i\mathbf{p} \cdot \mathbf{r}/\hbar)$ as usual [181].

Now let us derive from the second quantization the real space representations in first quantization using the results above. First consider a single-particle state $|\psi\rangle = |1_\psi\rangle$,

$$\begin{aligned} \langle \mathbf{r} | \psi \rangle &= \langle 0 | \hat{a}_{\mathbf{r}} \hat{a}_\psi^\dagger | 0 \rangle = \langle 0 | \hat{a}_{\mathbf{r}} \left(\int d\mathbf{r}' \psi(\mathbf{r}') \hat{a}_{\mathbf{r}'}^\dagger \right) | 0 \rangle \\ &= \int d\mathbf{r}' \psi(\mathbf{r}') \langle 0 | \hat{a}_{\mathbf{r}} \hat{a}_{\mathbf{r}'}^\dagger | 0 \rangle = \int d\mathbf{r}' \psi(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') = \psi(\mathbf{r}), \end{aligned} \quad (2.28)$$

which is just the usual definition. Since there is just one particle, there is no symmetry bookkeeping. Next, consider the $i \neq j$ two-particle state $|\lambda_i, \lambda_j\rangle = |1_{\lambda_i}, 1_{\lambda_j}\rangle$,

$$\begin{aligned} \langle \mathbf{r}_1, \mathbf{r}_2 | \lambda_i, \lambda_j \rangle &= \langle 0 | \frac{\hat{a}_{\mathbf{r}_1} \hat{a}_{\mathbf{r}_2}}{\sqrt{2}} \hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_j}^\dagger | 0 \rangle \\ &= \frac{1}{\sqrt{2}} \langle 0 | \hat{a}_{\mathbf{r}_1} \hat{a}_{\mathbf{r}_2} \left(\int d\mathbf{r}'_1 d\mathbf{r}'_2 \lambda_i(\mathbf{r}'_1) \lambda_j(\mathbf{r}'_2) \hat{a}_{\mathbf{r}'_1}^\dagger \hat{a}_{\mathbf{r}'_2}^\dagger \right) | 0 \rangle \\ &= \frac{1}{\sqrt{2}} \int d\mathbf{r}'_1 d\mathbf{r}'_2 \lambda_i(\mathbf{r}'_1) \lambda_j(\mathbf{r}'_2) \left(\langle 0 | \hat{a}_{\mathbf{r}_1} \hat{a}_{\mathbf{r}_2} \hat{a}_{\mathbf{r}'_1}^\dagger \hat{a}_{\mathbf{r}'_2}^\dagger | 0 \rangle \right) \\ &= \frac{1}{\sqrt{2}} \int d\mathbf{r}'_1 d\mathbf{r}'_2 \lambda_i(\mathbf{r}'_1) \lambda_j(\mathbf{r}'_2) \\ &\quad \times (\delta(\mathbf{r}_1 - \mathbf{r}'_1) \delta(\mathbf{r}_2 - \mathbf{r}'_2) + \delta(\mathbf{r}_1 - \mathbf{r}'_2) \delta(\mathbf{r}_2 - \mathbf{r}'_1)) \\ &= \frac{1}{\sqrt{2}} (\lambda_i(\mathbf{r}_1) \lambda_j(\mathbf{r}_2) + \lambda_i(\mathbf{r}_2) \lambda_j(\mathbf{r}_1)) = \Psi_{\lambda_i, \lambda_j}^+(\mathbf{r}_1, \mathbf{r}_2), \end{aligned} \quad (2.29)$$

which is seen to be on the appropriate symmetrized form of Eq. (2.13). Consider finally the $i = j$ two-particle state $|2_{\lambda_i}\rangle$,

$$\begin{aligned} \langle \mathbf{r}_1, \mathbf{r}_2 | 2_{\lambda_i} \rangle &= \langle 0 | \frac{\hat{a}_{\mathbf{r}_1} \hat{a}_{\mathbf{r}_2}}{\sqrt{2}} \frac{\hat{a}_{\lambda_i}^\dagger \hat{a}_{\lambda_i}^\dagger}{\sqrt{2}} | 0 \rangle = \frac{1}{2} (\lambda_i(\mathbf{r}_1) \lambda_i(\mathbf{r}_2) + \lambda_i(\mathbf{r}_2) \lambda_i(\mathbf{r}_1)) \\ &= \lambda_i(\mathbf{r}_1) \lambda_i(\mathbf{r}_2), \end{aligned} \quad (2.30)$$

which also has the right normalization and symmetry.

In conclusion, the second quantization is a particularly effective framework for dealing with multi-particle quantum states.

2.3.3 Many-body Hamiltonians

The general N -many-body Hamiltonian without spin in the real space coordinate representation with two-body interactions reads, in first quantization,

$$\hat{H}_{\text{mb}}^{(N)} = \sum_i \hat{H}_0(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \hat{U}_{\text{int}}(\mathbf{r}_i, \mathbf{r}_j). \quad (2.31)$$

The factor $1/2$ adjusts for double counting of pairs and the single-particle Hamiltonian is the usual

$$\hat{H}_0 = \hat{T} + \hat{U}_{\text{ext}}(\mathbf{r}) = \frac{\hat{\mathbf{p}}^2}{2m} + \hat{U}_{\text{ext}}(\mathbf{r}) = -\frac{\hbar^2}{2m} \nabla_{\mathbf{r}}^2 + U_{\text{ext}}(\mathbf{r}), \quad (2.32)$$

where $\nabla_{\mathbf{r}}$ is the gradient and $\hat{U}_{\text{ext}}$ is some external trapping potential operator (the last equality is explicitly in real space representation). In this thesis, the primary platforms of interest are trapped indistinguishable bosonic atoms [86, 92, 158, 178] that are,

- (1) neutrally charged
- (2) ultracold
- (3) in the dilute limit, $\langle n \rangle a^3 \ll 1$, where $\langle n \rangle$ is the mean density of the atomic gas and a the scattering length (defined momentarily). This can also be thought of as a weakly interacting limit.

Let us motivate why Eqs. (2.31)-(2.32) are valid descriptions thereof.

As a consequence of (1), i.e. the absence of long-range Coulomb forces, the atoms experience only short-range van der Waals-like interaction. The exact microscopic details of such scattering processes are notoriously complicated, but, as it turns out, reduce to a surprisingly simple coincidental description due to (2). Briefly, solving scattering problems [181, 182] for sufficiently short-ranged and spherically symmetric interactions entail analyzing the asymptotic behavior of the scattered wave function expanded in terms of so-called partial waves. Each partial wave is associated with one of usual angular momentum quantum numbers $l = 0, 1, 2, \dots = s, p, d, \dots$ — this leads to the definition of species dependent scattering lengths, a_l , and in the limit of low energies (2) it can be shown that a_s is the only dominant contribution for bosons⁸. For rubidium 87 it is typically on orders

⁸ Scattering between identical bosons [183] (fermions) have vanishing odd $l = 1, 3, \dots$ (even $l = 0, 2, \dots$) contributions. Identical fermions scattering is therefore instead dominated by a_p in the low energy regime. In general, the symmetry requirements have strong consequences for quantum interferences [184, 185].

of tens or hundreds of Bohr radii [158, 186], i.e. roughly orders of nm, and with the specific value having some variance depending on the internal states of the collision pair. It can then be shown that the interaction can be properly replaced by a pseudo-potential that yields the correct scattering phases,

$$U_{\text{int}}(\mathbf{r}_i, \mathbf{r}_j) \rightarrow U_{\text{int}}(\mathbf{r}_i - \mathbf{r}_j) = g_{3\text{D}} \delta(\mathbf{r}_i - \mathbf{r}_j) = \frac{4\pi\hbar^2 a_s}{m} \delta(\mathbf{r}_i - \mathbf{r}_j), \quad (2.33)$$

which is described entirely in terms of the s -wave scattering length and a true δ -contact potential⁹ in $d = 3$ dimensional space. Here we also defined the coupling strength $g_{3\text{D}} = 4\pi\hbar^2 a_s / m$ which is evidently very weak (compare e.g. numbers in Ref. [2]). Lastly, the short range (1) and the diluteness condition (3) ensure that only two-body interactions need be included in Eq. (2.31) to a good approximation — three-body or higher processes are strongly suppressed. Writing out the first few N we then have

$$\hat{H}_{\text{mb}}^{(1)} = \hat{H}_0(\mathbf{r}) = \hat{T} + \hat{U}_{\text{ext}}(\mathbf{r}), \quad (2.34)$$

$$\hat{H}_{\text{mb}}^{(2)} = \hat{H}_0(\mathbf{r}_1) + \hat{H}_0(\mathbf{r}_2) + g_{3\text{D}} \delta(\mathbf{r}_1 - \mathbf{r}_2), \quad (2.35)$$

$\vdots$

As with quantum states, all operators can also be mapped to the second quantization formulation and any one of these [179, 180] can be written in terms of the fundamental creation and annihilation operators for the single-particle modes. For example, the many-body Hamiltonian Eq. (2.31) with contact interactions can be written as

$$\hat{H}_{\text{mb}} = \int \hat{a}_{\mathbf{r}}^\dagger \hat{H}_0 \hat{a}_{\mathbf{r}} d^d \mathbf{r} + \frac{g_{3\text{D}}}{2} \int \hat{a}_{\mathbf{r}}^\dagger \hat{a}_{\mathbf{r}}^\dagger \hat{a}_{\mathbf{r}} \hat{a}_{\mathbf{r}} d^d \mathbf{r}. \quad (2.36)$$

This is the second quantized equivalent of the spatial basis representation — recall the operator definition $\hat{a}_{\mathbf{r}}^\dagger |0\rangle = |\mathbf{r}\rangle$. That is, $\hat{a}_{\mathbf{r}}^\dagger$ ($\hat{a}_{\mathbf{r}}$) conjures into (dispels from) existence a particle at location $\mathbf{r}$ and is often referred to as the *field operator*¹⁰. Immediately we notice that, as opposed to the first quantized version, we need not specify the number of particles ^(N) due to the wonders of the formalism. A few short energy calculations will make this operator assignment plausible. The results are exactly what we would get from first quantization, but there we have to fuss over the symmetrization bookkeeping.

⁹ Strictly speaking there should be a regularizing derivative after the δ , but this can be omitted if the acted-on function is regular at zero [86].

¹⁰ Sometimes the field operator is denoted $\hat{\Psi}(\mathbf{r})$ but I personally find this departure from otherwise consistent notation unnecessarily confusing.

First, let us transform the first term into the eigenbasis defined by $\hat{H}_0 |E_i\rangle = E_i |E_i\rangle$ with real space representation $\epsilon_i(\mathbf{r}) = \langle \mathbf{r} | E_i \rangle$ using Eqs. (2.27),

$$\begin{aligned} \int \hat{a}_{\mathbf{r}}^\dagger \hat{H}_0 \hat{a}_{\mathbf{r}} d^d \mathbf{r} &= \int \left(\sum_i \hat{a}_{E_i}^\dagger \epsilon_i^*(\mathbf{r}) \right) \hat{H}_0 \left(\sum_j \hat{a}_{E_j} \epsilon_j(\mathbf{r}) \right) d^d \mathbf{r} \\ &= \sum_{i,j} \hat{a}_{E_i}^\dagger \left(\int \epsilon_i^*(\mathbf{r}) \hat{H}_0 \epsilon_j(\mathbf{r}) d^d \mathbf{r} \right) \hat{a}_{E_j} = \sum_{i,j} E_i \langle E_i | E_j \rangle \hat{a}_{E_i}^\dagger \hat{a}_{E_j} \\ &= \sum_i E_i \hat{n}_{E_i}. \end{aligned} \quad (2.37)$$

It is then trivial to measure the energy of a multi-particle configuration distributed in any of the single-particle eigenstates $|n_{E_0}, n_{E_1}, \dots\rangle$,

$$\left\langle n_{E_0}, n_{E_1}, \dots \left| \int \hat{a}_{\mathbf{r}}^\dagger \hat{H}_0 \hat{a}_{\mathbf{r}} d^d \mathbf{r} \right| n_{E_0}, n_{E_1}, \dots \right\rangle = \sum_i E_i n_{E_i}. \quad (2.38)$$

This shows that the missing $\langle N \rangle$ on the Hamiltonian is justified since it is instead now the states that carry that information ($N = \langle \psi | \hat{N} | \psi \rangle$ where $\hat{N} = \sum \hat{n}_{\lambda_i}$). A strength of the formalism is that any single-particle quantity is readily calculated — simply transform into the respective diagonal basis, count the number of particles in each mode $|\lambda_i\rangle$, and weigh it by the quantity with associated quantum number λ_i . Of course we still need λ_i and sometimes $\Lambda_i(\mathbf{r}) = \langle \mathbf{r} | \lambda_i \rangle$ to be readily calculable in the first place, and this is typically achieved by numerical means. Similarly, an infinitesimal time evolution $\hat{\mathcal{U}} \approx 1 - i\hat{H}_0 \delta t$ with this Hamiltonian ascribes a phase,

$$|n_{E_0}, n_{E_1}, \dots\rangle \rightarrow \left(1 - i \left(\sum_i E_i n_{E_i} \right) \delta t / \hbar \right) |n_{E_0}, n_{E_1}, \dots\rangle, \quad (2.39)$$

and the multi-particle state formed from single-particle eigenstates of H_0 is also stationary with respect hereto.

Now let us consider the second term and likewise transform into the eigenbasis for $\hat{H}_0$,

$$\begin{aligned} \frac{g_{3D}}{2} \int \hat{a}_{\mathbf{r}}^\dagger \hat{a}_{\mathbf{r}}^\dagger \hat{a}_{\mathbf{r}} \hat{a}_{\mathbf{r}} d^d \mathbf{r} &= \frac{1}{2} \sum_{ijkl} \left(g_{3D} \int \epsilon_i^* \epsilon_j^* \epsilon_k \epsilon_l d^d \mathbf{r} \right) \hat{a}_{E_i}^\dagger \hat{a}_{E_j}^\dagger \hat{a}_{E_k} \hat{a}_{E_l} \\ &\equiv \frac{1}{2} \sum_{ijkl} U_{ijkl} \hat{a}_{E_i}^\dagger \hat{a}_{E_j}^\dagger \hat{a}_{E_k} \hat{a}_{E_l}, \end{aligned} \quad (2.40)$$

where U_{ijkl} (the parenthesis content) is the interaction strength between the modes $ijkl$. For simplicity let us measure the energy of $|1_{E_n}, 1_{E_m}\rangle$,

$$\begin{aligned} &\frac{1}{2} \sum_{ijkl} U_{ijkl} \langle 1_{E_n}, 1_{E_m} | \hat{a}_{E_i}^\dagger \hat{a}_{E_j}^\dagger \hat{a}_{E_k} \hat{a}_{E_l} | 1_{E_n}, 1_{E_m} \rangle \\ &= \frac{1}{2} \sum_{ijkl} U_{ijkl} \langle 0|0\rangle (\delta_{i,n} \delta_{j,m} \delta_{k,n} \delta_{l,m} + \delta_{i,m} \delta_{j,n} \delta_{k,m} \delta_{l,n}) = \frac{U_{nnmm} + U_{mnmn}}{2} \\ &= g_{3D} \int |\epsilon_n(\mathbf{r})|^2 |\epsilon_m(\mathbf{r})|^2 d^d \mathbf{r}. \end{aligned} \quad (2.41)$$

That is, the collisional interaction element between two particles distributed in single-particle eigenstates is given by their density overlaps and the coupling constant. Notice in general, however, that the operator $\hat{a}_{E_i}^\dagger \hat{a}_{E_j}^\dagger \hat{a}_{E_k} \hat{a}_{E_l}$ is very clearly responsible, in this case, for inducing transitions $kl \rightarrow ij$ with strength U_{ijkl} . Upon infinitesimal time evolution with this Hamiltonian we have

$$|n_{E_0}, n_{E_1}, \dots\rangle \rightarrow |n_{E_0}, n_{E_1}, \dots\rangle - i\delta t/\hbar \times \sum_{ijkl} U_{ijkl} |n_{E_i} + 1, n_{E_j} + 1, n_{E_k} - 1, n_{E_l} - 1\rangle |\dots\rangle, \quad (2.42)$$

where $|\dots\rangle$ are the unaffected modes $q \neq i, j, k, l$. The single-particle eigenenergy modes are thus not stationary states for the interaction and will be mixed in upon time evolution.

We have now seen what many-body Hamiltonians look like in both first and second quantization. In both instances, the strategy for application is always to describe the system in some appropriate basis which may or may not be the single-particle energy eigenstates.

2.3.4 Bose-Einstein Condensates

Let us now consider a macroscopic gas of bosons ($N \gg 1$) residing in a trapping geometry.

Non-interacting gas — exact ground state

We consider first a non-interacting gas ($g_{3D} = 0$) in which case the single-particle energy eigenstates $\hat{H}_0(\mathbf{r}) |E_i\rangle = E_i |E_i\rangle$ (assumed non-degenerate for simplicity, $E_i < E_{i+1}$) are also many-body energy eigenstates cf. Eq. (2.38). From the grand canonical ensemble we obtain the bosonic distribution into the these modes [158]

$$N = \sum_i \langle \hat{n}_i \rangle = \sum_i \left(\exp \left(\frac{E_i - \mu}{k_B T} \right) - 1 \right)^{-1}, \quad (2.43)$$

where the μ is the chemical potential, T is the temperature, and k_B is the Boltzmann constant. For physically meaningful results we see that $E_0 - \mu \geq 0 \Rightarrow \mu \leq E_0$. For a fixed N and T , the expression above implicitly defines $\mu(N, T)$. As $\mu(N, T \rightarrow \infty) \rightarrow -\infty$, we regain in the high T limit a classical Maxwell-Boltzmann distribution $\langle n_i \rangle \approx \exp((E_i - \mu)/(k_B T))^{-1} \ll 1$ where the boson particle statistics are irrelevant. As $T \rightarrow 0$, however, the population in the ground state begins piling up because the denominator in $\langle \hat{n}_i \rangle$ diverges and below some critical temperature, T_c , it becomes macroscopically occupied in the thermodynamic limit,

$$\lim_{\substack{N \rightarrow \infty \\ T \leq T_c}} \frac{\langle \hat{n}_0 \rangle}{N} > 0. \quad (2.44)$$

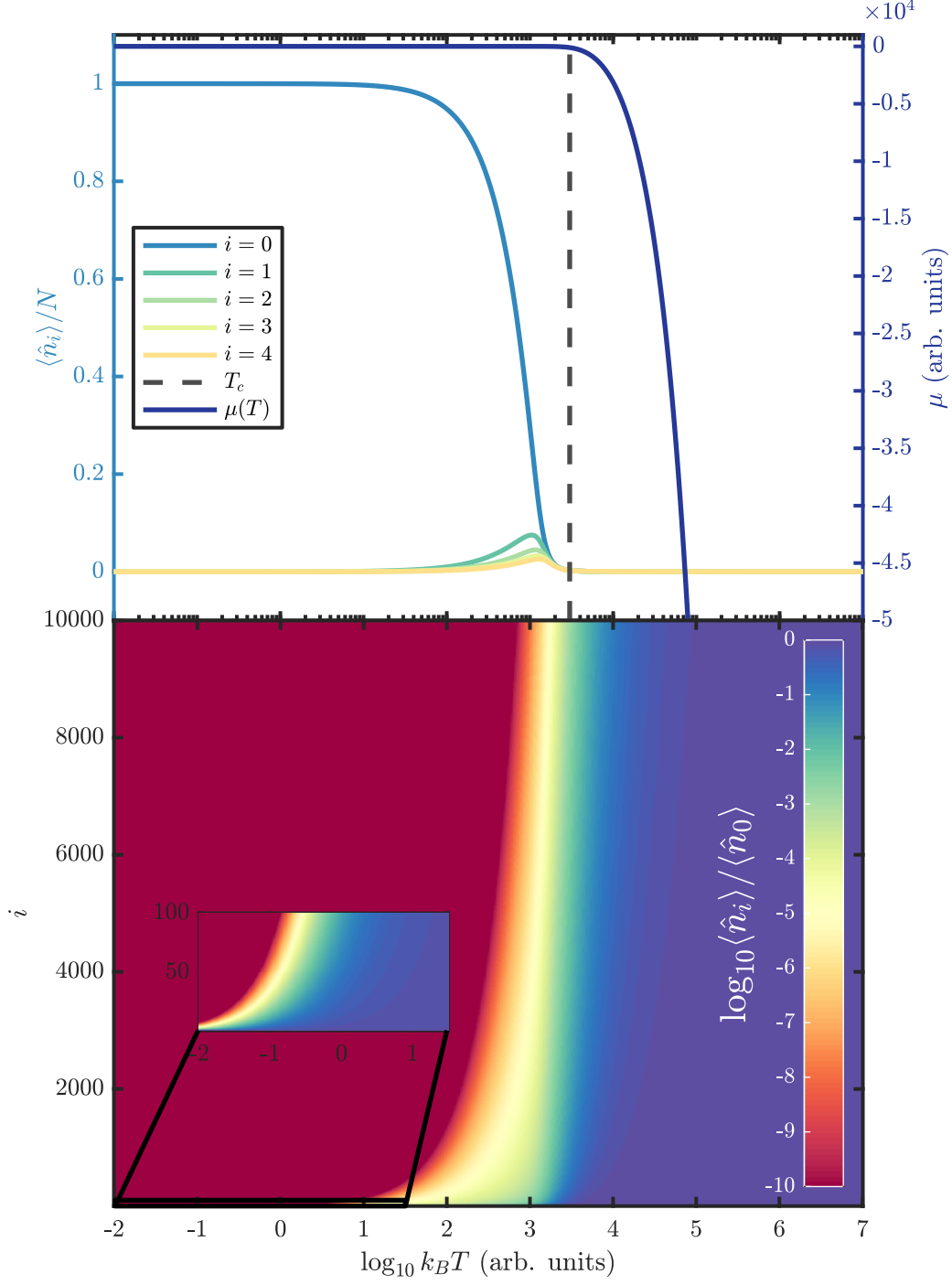


Figure 2.3: Example of Bose-Einstein condensation for $N = 10^4$ interaction-free atoms with a non-degenerate harmonic spectrum $E_i = \hbar\omega(\frac{1}{2} + i)$. The results were numerically obtained by solving Eq. 2.43 in simulation units where $\hbar = \omega = 1$. Top: The N -normalized state occupation in lowest single-particle modes $|E_i\rangle$ and the chemical potential as a function of temperature T . The vertical dashed line roughly (chosen manually) indicates T_c where the condensation sets in. Bottom: The $\langle \hat{n}_0 \rangle$ -normalized state occupation in all single-particle modes $|E_i\rangle$. As the temperature decreases below T_c , the thermal state population drastically tightens and the inset shows that this continues until all particles are condensed in the ground state. The color scale cuts off at -10 for readability but the actual values approach -300 .

This is the signature of Bose-Einstein condensation [158] and at $T = 0$ we have exactly $|n_{E_0}, n_{E_1}, n_{E_2}, \dots\rangle = |N, 0, 0, \dots\rangle = |N\rangle |0\rangle$. These ideas are numerically illustrated in Fig. 2.3. In the fully condensed case, the exact ground state in first quantization is the properly symmetrized wave function

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \prod_{i=1}^N \epsilon_0(\mathbf{r}_i), \quad (2.45)$$

where $\epsilon_0(\mathbf{r}_i) = \langle \mathbf{r}_i | E_0 \rangle$. For once there is no tedious bookkeeping because all particles are in the same state. Since there are no interactions, each particle enjoys only its own company and its characteristics and subsequent dynamics are given by N identical copies of the usual Schrödinger equations

$$\hat{H}_0 \epsilon_0 = E_0 \epsilon_0, \quad i\hbar \frac{\partial \lambda_1}{\partial t} = \hat{H}_0 \lambda_1, \quad (2.46)$$

where $\lambda_1(\mathbf{r}, t = 0) = \epsilon_0(\mathbf{r})$ such that $|n_{\lambda_1(t)}, 0, 0, \dots\rangle$ remains condensed.

Interacting gas — approximate ground state

The formation of BEC is a purely quantum statistical effect which itself is due to the symmetry requirements and occurs even without interactions. Nevertheless, we know that short-range binary collisions are relevant in our affairs and this slightly complicates the situation. We can equally well proceed with its inclusion in either the first or second quantization and we somewhat arbitrarily choose the former here. The crux of the matter is that the single-particle energy eigenstates are not eigenstates for the full many-body Hamiltonian in the interacting case ($g_{3D} \neq 0$) cf. Eq. (2.42). We therefore expect the interactions to modify the “natural” single-particle basis wave functions. Let us imagine this new and currently unknown (variational) basis by $|\tilde{E}_i\rangle$, ascribed with “ $\sim$ ”. We shall only require that they coincide with the single-particle eigenstates in the limit $g_{3D} \rightarrow 0$ and when the interactions are finite they will generally be related/mixed by a basis transformation similar to Eq. (2.27). In the lowest order of approximation, suppose the condensation proceeds as before, but now into $|n_{\tilde{E}_0}\rangle = |N\rangle$ where we now omit writing out all the empty modes. The corresponding properly symmetrized wave function in first quantization is then

$$\tilde{\Psi}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \prod_{i=1}^N \tilde{\epsilon}_0(\mathbf{r}_i), \quad (2.47)$$

where $\tilde{\epsilon}_0(\mathbf{r}_i) = \langle \mathbf{r}_i | \tilde{E}_0 \rangle$ with appropriate normalization $\int \tilde{\epsilon}_0 d^d \mathbf{r} = 1$. Although this looks similar to Eq. 2.45, it is important to note that this form is just a convenient ansatz appropriate for capturing the

essential features of the condensation. That is, we can self-consistently determine $\tilde{\epsilon}_0$ by minimizing the energy,

$$\tilde{\epsilon}_0 = \arg \min_{\tilde{\epsilon}_0} \left\langle \tilde{\Psi} \left| H_{\text{mb}}^{(N)} \right| \tilde{\Psi} \right\rangle - \mu (\langle \tilde{\epsilon}_0 | \tilde{\epsilon}_0 \rangle - 1), \quad (2.48)$$

where the second term fixes the number of particles (recall $\mu = \mu(N)$) and ensures normalization. This type of minimization can be achieved by, e.g., Gâteaux derivatives (cf. Sec. 4.3) but we do not expand on the details here¹¹. The final result is an eigenequation that, in the spatial representation, reads

$$\hat{H}_{\text{gp}}(\mathbf{r}, \psi) \psi \equiv (H_0(\mathbf{r}) + g_{3\text{D}} |\psi(\mathbf{r})|^2) \psi = \mu \psi, \quad (2.49)$$

where $\hat{H}_{\text{gp}}$ is the Gross-Pitaevskii (GP) Hamiltonian and we have defined the condensate wave function $\psi(\mathbf{r}) = \sqrt{N} \tilde{\epsilon}_0(\mathbf{r})$ which evidently is normalized to N . From this it is also natural to define the condensate density $n(\mathbf{r}) = |\psi(\mathbf{r})|^2$. It also turns out that higher-order modes $i > 0$ of the stationary GP equation are physically meaningful.

It is quite remarkable how this inherently N -many-body system in the ansatz of Eq. (2.47) yields a stationary single-particle-like Schrödinger equation where the chemical potential plays the role of the usual eigenenergy, characterized by a *self-interaction* term, $g_{3\text{D}} |\psi|^2$. It is easy to see that the self-interaction acts as an additional spatial potential and therefore tends to locally broaden (sharpen) the condensate wave function for repulsive (attractive) $g_{3\text{D}}$ when compared to the interaction-free single-particle state — this is exactly what one would qualitatively expect (see also Ch. 5 for a quantitative verification). It can also be shown that the condensate dynamics is given by [158]

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}_{\text{gp}} \psi = (H_0(\mathbf{r}) + g_{3\text{D}} |\psi(\mathbf{r})|^2) \psi, \quad (2.50)$$

in complete analogy to the usual time dependent Schrödinger equation. These celebrated GP equations are also sometimes called nonlinear Schrödinger equations¹².

It is important to note that even upon time evolution, the atoms are assumed to remain fully condensed and from a computational perspective it is “as if” $N = 1$ (with a few caveats pertaining to the self-interaction described in Ch. 5). Thus, although this modeling is often a sufficient or even excellent description for $N \gg 1$ trapped bosons, it lacks the capacity for describing fine-grained many-body particle correlations as elucidated in the following.

¹¹ The assumption of the state residing in a trapping geometry such that $\tilde{\epsilon}_0(\pm\infty) = 0$ is utilized in these calculations.

¹² Incidentally, such nonlinear equations also appear in optics and this lead to the predictions of bright solitons in attractive boson gases [187].

The meanfield and beyond

The so-called *meanfield*¹³ description [149] by $\psi(\mathbf{r})$ is clearly a *much* more tangible object than the monolithic $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ and allows computation of several important static, dynamic, and thermodynamic quantities [158]. It also plays the role of an order parameter [159] for the condensation phase transition — recall $\psi \sim \sqrt{n_0}$ which is vanishing for $T > T_c$ and finite for $T < T_c$ as seen in Fig. 2.3. Nevertheless, the single-mode meanfield is only the simplest possible description of BEC phenomena. There are many, many extensions and further considerations that are relevant for describing the behavior of ultracold gases/atoms [158]. For example, the meanfield cannot include entanglement and correlations because it represents a product form by construction cf. Eq. (2.47) which was the lowest order approximation possible. First order quantum corrections to the fully condensed state can be achieved in the *Bogoliubov* approximation where the condensed atoms are treated as a classical field and the remaining modes as quantum fluctuations, i.e. the second quantized field operator is approximately $\hat{a}_{\mathbf{r}} \sim \psi(\mathbf{r}) + \delta\hat{a}_{\mathbf{r}}$. Further, experiments are never performed truly at $T = 0$ and this gives rise to additional condensate depletion into a thermal contribution widely spread throughout the $i > 0$ modes, though non-macroscopically in each i . There may also be other types of interactions present that generalizes the meanfield and associated the GP equations, e.g., long-range dipolar interactions [167, 188] which introduces terms proportional to $|\psi|^3$ and $|\psi|^4$.

Lastly, it is also entirely possible to macroscopically populate several modes simultaneously and this is referred to as fragmented condensates. Such observations could already be gleaned from Fig. 2.3 around T_c where the lowest excited states momentarily became somewhat macroscopically occupied. Theoretically describing fragmented [158, 187, 189] condensates is much more complicated and is typically numerically treated [190] within the ansatz known as *multiconfigurational time-dependent Hartree method for bosons* (MCTDHB) [191–196] which provides a natural M -multi-mode description. For $M = 1$ one incidentally recovers the GP meanfield and we already know that this cannot account for truly many-body processes — see e.g. Ref. [187] for a numerical example displaying profound differences. In fact, further to this point, MCTDHB (and its cousins MCTDH and MCTDHF) is an extremely powerful tool for describing many-body systems. A brief motivation and description thereof is as follows.

Recall that one of the first assumptions made in this section was that the single-particle states $\{|\lambda_i\rangle\}$ and their second quantized represen-

¹³ The origin of this name is that an individual particle has a vanishing contribution (it could be removed with impunity) to the goings-on when compared to the contribution due to the macroscopic ensemble which is composed of the $N - 1$ other particles. Since this argument can be made for each particle, this implies that a description of a *single* particle living in a self-consistent *meanfield* due to the rest is meaningful.

tations $\{|n_{\lambda_i}\rangle\}$ were time independent up until Eqs. (2.46) and (2.50), respectively, which then described the dynamics of a single mode occupied by all N atoms at all times, $|n_{\lambda_1(t)}\rangle = |N\rangle$. The MCTDHB ansatz is the natural generalization to this,

$$|\Psi(t)\rangle = \sum_{\substack{n_{\lambda_1}, \dots, n_{\lambda_M} \\ N = \sum_i n_{\lambda_i}}} c_{\mathbf{n}}(t) |\mathbf{n}(t)\rangle \quad (2.51)$$

where $c_{\mathbf{n}}(t)$ are expansion coefficients for the M modes given by $|\mathbf{n}(t)\rangle = |n_{\lambda_1(t)}, \dots, n_{\lambda_M(t)}\rangle$. Note that this is essentially the same as Eq. (2.14), with the important distinctions that (1) only a finite number of M modes are considered, and (2) both the coefficients and modes are time dependent. It is actually not at all difficult to motivate this expansion¹⁴. The idea is to find the *best possible* time dependent set of M coefficients and single-particle states, $\{c_{\mathbf{n}}(t)\}, \{\Lambda_i(\mathbf{r}, t)\}$, to describe $|\Psi(t)\rangle$ by variationally minimizing [192]

$$S[\{c_{\mathbf{n}}(t)\}, \{\Lambda_i(\mathbf{r}, t)\}] = \int \left(\left\langle \Psi \left| \hat{H} - i \frac{\partial}{\partial t} \right| \Psi \right\rangle - \sum_{k,j}^M \mu_{k,j}(t) (\langle \lambda_k | \lambda_j \rangle - \delta_{k,j}) \right). \quad (2.52)$$

The variational derivatives of S then yields coupled equations of motion for the single-particle wave functions and their occupation numbers. This approach obviously transcends the study of BEC and can be used in any setting. In a much similar way, it can be used to describe phenomena which are not captured [193] by the otherwise celebrated lowest band Bose-Hubbard model described later in this chapter. We shall nevertheless not concern ourselves further with the MCTDHB methods in this thesis.

¹⁴ Suppose we had a earth laboratory with tightly confining potential V centered on coordinates $\mathbf{r}_{\text{earth}}$ and a laboratory somewhere on the moon with an identical potential centered on $\mathbf{r}_{\text{moon}}$. Obviously, the corresponding energy eigenstates $\{|E_{i,\text{earth}}\rangle\}$ and $\{|E_{i,\text{moon}}\rangle\}$ are individually complete bases. Suppose then at time $t = 0$ we procure an atom in the earth lab described by $\psi(\mathbf{r}) = \epsilon_{0,\text{earth}}(\mathbf{r} - \delta\mathbf{r})$, i.e. one that is microscopically displaced from the earth ground state. We can then write $|\psi\rangle = \sum_i^{M_{\text{earth}}} c_i |E_{i,\text{earth}}\rangle = \sum_i^{M_{\text{moon}}} c'_i |E_{i,\text{moon}}\rangle$ equally well — but which expansion would we prefer? Clearly the former, because we would need “infinitely” fewer expansion coefficients to gain an accurate description, $M_{\text{earth}} \ll M_{\text{moon}} \rightarrow \infty$. Suppose then that the atom was coherently transported in time T to the moon lab such that $|\psi(\mathbf{r}, T)\rangle = \epsilon_{0,\text{moon}}(\mathbf{r} - \delta\mathbf{r})$. Now clearly $M_{\text{moon}} \ll M_{\text{earth}} \rightarrow \infty$ yields a good description and the moon expansion is preferable. However, for intermediate times $0 \ll t \ll T$ neither of the expansions are feasible since both $M_{\text{earth}}, M_{\text{moon}} \rightarrow \infty$. Instead, it would be a much better to imagine an approach where the basis functions *themselves* are time dependent such that we only ever need up to some finite number M of them for an accurate description.

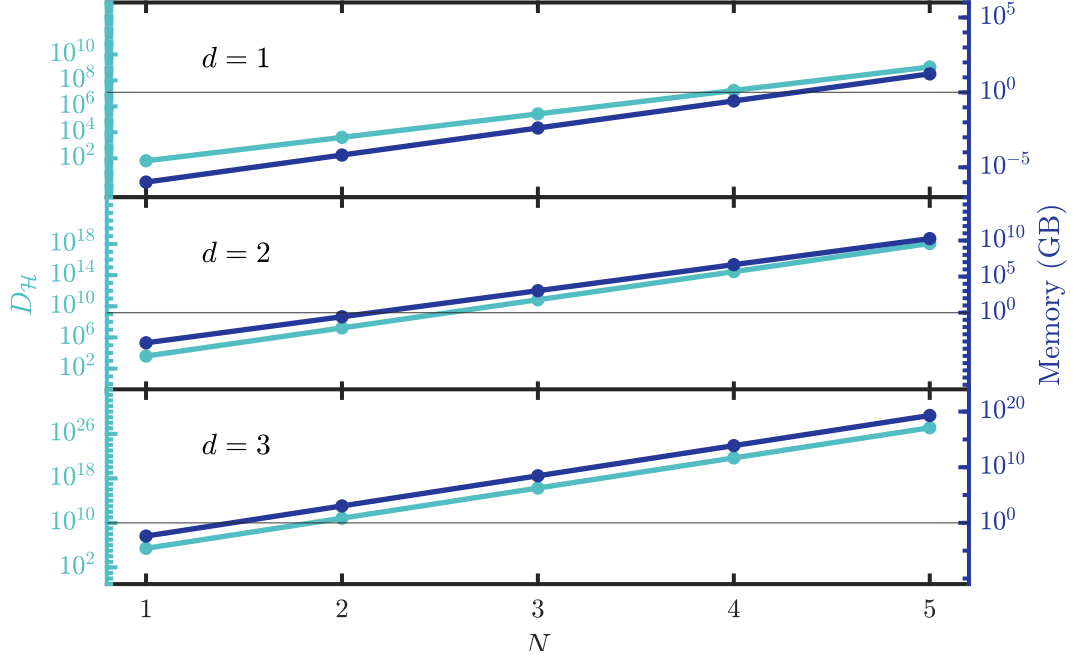


Figure 2.4: Scaling of Hilbert space D_H and the memory requirements to store a generic wave function as a function of N for different d in first quantization. Here we assume a relatively modest spatial discretization of 64 grid points in all d dimensions.

2.4 PHYSICAL MODELING IN 1D

The developments in the previous section kept, to the broadest extent possible, the real space dimensionality d variable. Inevitably, however, we must come to terms with the likely fact that $d = 3$ is the correct description of Nature. This imposes serious concerns for the (exact) numerical study of N -body systems (with Nd coordinates/degrees of freedom) because, due to the superposition principles of quantum mechanics, we must consider in simultaneity (1) the amplitudes of all possible state configurations, and (2) the matrix representation of relevant operators in some appropriate bases. In first quantization, this means that even $N = 1, d = 3$ is a computationally demanding, yet still ultimately surmountable [197], task for spatial discretization of sufficiently fine resolution. Yet, the situation for $N = 2, d = 3$ already becomes intractable as shown in Fig. 2.4 and this scenario can scarcely be called “many-body” to begin with. Employing second quantization alleviates these pains somewhat (it circumvents symmetry bookkeeping and usually involves storing coefficients of appropriate an single-particle basis expansion rather than for each spatial coordinate) but even in this case we must evaluate expensive $N = 1, d = 3$ integrals such as in Eq. (2.40) and the exponential scaling in N is under all circumstances inevitable.

2.4.1 Separation of variables

To make headway towards more realistic computational treatments we notice the scaling is less severe, albeit still quite limiting, when $d < 3$. That is, we can extend the feasible reach of our numerics if we can somehow perform a *dimensionality reduction* that allows us to “ignore”, at least approximately, certain degrees of freedom.

Spin

In a certain sense, we have already performed a dimensionality reduction in not including particle spin. So long as there are no significant spin-dependent interactions, and thus no corresponding term in the model Hamiltonian, the spin states $|\chi\rangle$ are trivially unit diagonal/uncoupled. The time evolution, e.g., of a properly symmetrized bosonic state $|\Psi^\pm\rangle |\chi^\pm\rangle$ is then $\hat{\mathcal{U}}(|\Psi^\pm\rangle |\chi^\pm\rangle) = (\hat{\mathcal{U}} |\Psi^\pm\rangle) |\chi^\pm\rangle = |\Psi^\pm(t)\rangle |\chi^\pm\rangle$ — for the purposes of computations regarding the spatial degrees of freedom it is then as if the spin states do not exist and, e.g.,

$$\begin{aligned} \langle \Psi^\pm(t), \chi^\pm | \hat{O} | \Psi^\pm(t), \chi^\pm \rangle &= \langle \Psi^\pm(t) | \hat{O} | \Psi^\pm(t) \rangle \langle \chi^\pm | \chi^\pm \rangle \\ &= \langle \Psi^\pm(t) | \hat{O} | \Psi^\pm(t) \rangle \end{aligned} \quad (2.53)$$

where $\hat{O} = \hat{O}(\{r_i\})$. Thus, we can simply opt to not include the spin degrees of freedom explicitly in the numerical implementation to reduce the dimensionality. If the Hamiltonian, instead, contained only spin terms then we would have the mirrored situation where it is if that the spatial states do not exist and dynamics occur only in the spins states.

These examples are the simplest instances of *separation of variables*, the purpose of which is to decompose the system into more manageable and uncoupled problem spaces. This technique is exact insofar as the Hamiltonian is separable into $\hat{H} = \hat{H}_{\text{space}} + \hat{H}_{\text{spin}}$ with obvious exclusive variable dependences. Under these separability conditions, the time evolution factors rigorously [198] as $\hat{\mathcal{U}} = \hat{\mathcal{U}}_{\text{space}} \otimes \hat{\mathcal{U}}_{\text{spin}}$ and

$$\hat{\mathcal{U}}(|\Psi^\pm\rangle |\chi^\pm\rangle) = (\hat{\mathcal{U}}_{\text{space}} |\Psi^\pm\rangle)(\hat{\mathcal{U}}_{\text{spin}} |\chi^\pm\rangle) = |\Psi^\pm(t)\rangle |\chi^\pm(t)\rangle \quad (2.54)$$

with the same consequences as the above examples. That is, we can treat spin and space variables *independently*. If we added a co-dependent “mixing” term $\hat{H}'_{\text{space and spin}}$ then the separation is not (at least exactly) possible, $\hat{\mathcal{U}} \neq \hat{\mathcal{U}}_{\text{space}} \otimes \hat{\mathcal{U}}_{\text{spin}}$, and the state will become entangled upon time evolution (i.e. not remain factored in a single aggregate space-spin term).

Space

The variable separation technique is great news, because we may be able to achieve further dimensionality reductions when considering

the single-particle Hamiltonian $H_0(\mathbf{r}) = \hat{T}_x + \hat{T}_y + \hat{T}_z + \hat{U}_{\text{ext}}(x, y, z)$. It is then incumbent on the potential to separate appropriately, e.g., $U_{\text{ext}}(x, y, z) = V_x(x) + V_y(y) + V_z(z)$, in which case the single-particle Hamiltonian can be written

$$\hat{H}_0(\mathbf{r}) = \hat{H}_{0,x}(x) + \hat{H}_{0,y}(y) + \hat{H}_{0,z}(z), \quad (2.55)$$

where e.g. $\hat{H}_{0,x}(x) = \hat{T}_x + V_x(x)$. In complete equivalence to the spin separation, the time evolution $\hat{\mathcal{U}} = \hat{\mathcal{U}}_x \otimes \hat{\mathcal{U}}_y \otimes \hat{\mathcal{U}}_z$ of an initial state $|\psi\rangle = |\psi_x\rangle |\psi_y\rangle |\psi_z\rangle$ is then

$$\hat{\mathcal{U}} |\psi\rangle = (\hat{\mathcal{U}}_x |\psi_x\rangle)(\hat{\mathcal{U}}_y |\psi_y\rangle)(\hat{\mathcal{U}}_z |\psi_z\rangle) = |\psi_x(t)\rangle |\psi_y(t)\rangle |\psi_z(t)\rangle. \quad (2.56)$$

We can therefore consider 3 independent problems of size $D_{\mathcal{H}_q}$, $q = x, y, z$, instead of the original 1 problem of size $D_{\mathcal{H}_x} D_{\mathcal{H}_y} D_{\mathcal{H}_z}$. In many situations we are only interested in the goings-on in a single direction, say x , and we then have a bona fide $N, d = 1$ problem. Similarly, it was separation of variables that allowed the N independent and identical copies of Eq. (2.46) in discussing the non-interacting BEC.

There are still a few problems to consolidate, however. The first is that most potentials only separate approximately as $U_{\text{ext}}(x, y, z) \approx V_x(x) + V_y(y) + V_z(z)$. The extent of validity typically covers areas near potential minima where a harmonic approximation with oscillation frequencies $\omega_{x,y,z}$ (second order Taylor expansion) is appropriate, for example in the center of an optical tweezer or lattice site as in Fig. 2.1. The second problem is that upon considering the ultracold collisional interaction Eq. (2.33) the many-body Hamiltonian Eq. (2.31) immediately becomes inseparable regardless of U_{ext} since $U_{\text{int}} \sim \delta(\mathbf{r}_i - \mathbf{r}_j)$ includes six coordinates. The resolution to both these problems is typically based on one of the two following assumptions: (1) dynamics in the direction of interest occur on much faster time scales ($\omega_x \gg \omega_y, \omega_z$), or (2) dynamics outside the direction of interest is frozen out ($\omega_x \ll \omega_y, \omega_z$). Although clearly two very different situations, both lead to the same approximate form of evolution $|\psi(t)\rangle \approx |\psi_x(t)\rangle |\psi_y, \psi_z\rangle$, where $|\psi_x(t)\rangle$ is the numerically simulated direction and $|\psi_y, \psi_z\rangle$ is assumed constant up to a phase and typically in an approximate ground state (also possibly separable $|\psi_y, \psi_z\rangle = |\psi_y\rangle |\psi_z\rangle$). The assumptions can also be combined, $\omega_z \ll \omega_x \ll \omega_y$, to the same effect (see, e.g., Ref. [74]). Since U_{int} is typically comparatively very weak, the error made in these assumptions is most often dominated by the approximate separability of U_{ext} , see e.g. Ref. [2].

The final caveat lies in the coupling strength, $g_{3D} = 4\pi\hbar^2 a_s/m$, which ascertains $d = 3$. Incidentally, it was only upon including the interaction that we could no longer neglect details of d in the previous chapter. Reducing the interaction dimensionality is generally tricky

because it is inherently a three dimensional scattering process. It is nevertheless still possible to retain the same form [199] such that,

$$U_{\text{int}}(\mathbf{r}_i - \mathbf{r}_j) \rightarrow V_{\text{int}}(x_i - x_j) = g_{1\text{D}}\delta(x_i - x_j) = 2a_s\hbar\sqrt{\omega_y\omega_z}\delta(x_i - x_j), \quad (2.57)$$

where $\omega_{y(z)}^2(x) = \partial_{yy(zz)}U_{\text{ext}}(\mathbf{r})/m$ are the local harmonic frequencies in the $y(z)$ direction that may be x dependent (if they vary slowly over the region of interest they can be taken as constant). The slightly different front factor is due to the algebraic content of the frequencies¹⁵ and analytically deriving suitable expressions can be somewhat involved [2, 74]. Remnants of the other axes are therefore present even upon dimensionality reduction but we are now able to comfortably entertain $N = 2, d = 1$, an effective quasi-one-dimensional [200] description¹⁶, as in Ref. [2]. In the meanfield description of BEC the effective coupling typically scales with the number of condensed atoms, e.g. $g_{1\text{D}} \rightarrow G_{1\text{D}}(N) = g_{1\text{D}}(N) \cdot N$ as described in more detail later in Ch. 5.

2.5 SYSTEM MODELS CONSIDERED IN THE THESIS

We are now in a position to describe the models relevant to this thesis. All Hamiltonians follow Schrödinger-like equations. Nearly all of these are quasi-one-dimensional using the machinery described in detail above. Only one of them, the Bose-Hubbard model, requires additional introduction in any mentionable capacity. Spin systems have not been given a thorough exposition here, but these are relatively easy to describe and simulate especially when the system is small as in Ref. [3]. In general, it turns out that an appropriate time evolution is the Suzuki-Trotter (“split-step”) decomposition for reasons discussed later in Ch. 4.

Additionally, we broadly refer to *exact diagonalization* as methods where we decide on particular, fixed basis (spatial, energy eigenstates, etc.) and “naïvely” store all the expansion coefficients. This works well when the Hilbert space is not too large as we saw in discussing Fig. 2.4. In the many-body limit, however, such descriptions become exponentially untenable and we may instead consider alternatives. This is the topic of Ch. 7 where we describe so-called *matrix product states* which are appropriate for certain (quasi)-one-dimensional systems.

¹⁵ As an instructive example, consider Taylor expanding around a minimum of an optical lattice $V_y + V_z = -v(\cos^2(ky) + \cos^2(kz))$. The harmonic approximation yields frequencies $\omega = \omega_y = \omega_z = \sqrt{2vk^2/m} = \sqrt{2v\pi^2/(ma^2)}$ and thus $g_{1\text{D}} = 2a_s\hbar\omega = g_{3\text{D}} \cdot \sqrt{mv/(2\hbar^2a^2)}$ where a is the lattice site separation. An additional length^{-1} unit comes from the $\delta(x)$ function.

¹⁶ Two-particle problems can also sometimes be recast into relative and center of mass coordinates which can then be separated if the potential is e.g. harmonic [201–203] and then solved analytically but we do not consider these situations further since the potentials considered here are never truly harmonic in all three dimensions.

2.5.1 One, Two, and Meanfield-Many Particle Models

Recapitulating the results from throughout this chapter we have the one [1, 5, 6] and two-particle [1, 2] Hamiltonians

$$\hat{H}^{(1)} = \hat{H}_0(x) = \hat{T} + \hat{V}(x) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x), \quad (2.58)$$

$$\hat{H}^{(2)} = \hat{H}_0(x_1) + \hat{H}_0(x_2) + g_{1D}\delta(x_1 - x_2). \quad (2.59)$$

with unit state normalization and effective interactions g_{1D} and may possibly be a function of x . The BEC meanfield description [1, 5, 6] is given by the GP Hamiltonian

$$\hat{H}_{\text{gp}} = \hat{H}_0(x) + G_{1D}|\psi|^2, \quad (2.60)$$

with the condensate wave function normalized to N atoms and the effective interactions are often on the form $G_{1D}(N) = g_{1D}(N) \cdot N$. The time evolution of all these Hamiltonians are further amenable to Fourier transformations during the Trotter steps [1] and this in particular allows a feasible first quantization treatment.

2.5.2 The Bose-Hubbard Model — A Genuine Many-Body System

Here we first state the Bose-Hubbard model [4]

$$\hat{H}_{\text{BH}} = \hat{H}_x^J + \hat{H}^U = J_x \sum_{i=1}^{N_s-1} \hat{h}_{[i,i+1]}^{J_x} + \frac{U}{2} \sum_{i=1}^{N_s} \hat{h}_{[i]}^U, \quad (2.61a)$$

$$\hat{h}_{[i,i+1]}^{J_x} = -(\hat{a}_{i+1}^\dagger \hat{a}_i + \text{h.c.}), \quad \hat{h}_{[i]}^U = \hat{n}_i(\hat{n}_i - 1), \quad (2.61b)$$

describing a fixed number of N_{part} spinless interacting bosons for a quasi-1D finite-periodic potential of N_{sites} sites in absence of additional external fields. The model stems from the second quantized Hamiltonian Eq. (2.36) expanded on an appropriate localized basis with a few simplifying assumption specified in detail below. Briefly, the creation and annihilation operators are associated with the lowest band Wannier wave functions that are localized on each site i and J_x is homogeneous hopping (or tunneling) strength for the nearest-neighbor interaction $\hat{h}^{J_x}$ and U the strength of the on-site interaction $\hat{h}^U$.

This instance of the Bose-Hubbard model captures two different many-body phases that are characterized by its ground state. For unit filling $N = N_{\text{sites}} = N_{\text{part}}$ and in the thermodynamic limit, we

associate with $J_x \gg U$ the superfluid phase and with $J_x \ll U$ the Mott insulating phase with respective ground states

$$|\text{SF}\rangle \propto \prod_{j=1}^N \left(\sum_{i=1}^N \hat{a}_i^\dagger \right) |0\rangle \\ \sim |N, 0, \dots, 0\rangle + |N-1, 1, \dots, 0\rangle + \dots + |0, 0, \dots, N\rangle, \quad (2.62)$$

$$|\text{Mott}\rangle \propto \prod_{j=1}^N \hat{a}_j^\dagger |0\rangle \sim |1, 1, \dots, 1\rangle, \quad (2.63)$$

corresponding to completely delocalized and localized particles, respectively. The crossover and phase diagram can be mapped using a variational meanfield approach [86, 92, 149, 204] and this requires inclusion of the chemical potential. In three dimensions, the full phase diagram famously consists of integer Mott-insulator lobes that are smooth and in lower dimensions these become more jagged [204]. The physics surrounding quantum phases are in general quite interesting because of the inherent sensitivity around the critical points [205, 206].

This model can be realized in different physical platforms which dictates constitutive relations between the model parameters U , J_x and the experimental parameters. We consider the implementation¹⁷ in an optical lattice with rubidium 87 atoms and use the system parameters from Ref. [74]: the lattice recoil energy is $E_r = \hbar^2 \pi^2 / (2a_s m) \approx 2.03 \text{ kHz} \cdot \hbar$ where $m = 87 \text{ amu}$ and $a_s = 101a_0$, the wave length is $\lambda = 1064 \text{ nm}$ with lattice spacing $a_{\text{lat}} = \lambda/2 = 532 \text{ nm}$, and fixed transverse trapping depths $v_y = v_z = 20 E_r$ (see manuscript [4] for further details). The constitutive relations are then derived as follows (see also Refs. [89, 207]).

Derivation of the Bose-Hubbard Model

Recounting the earlier discussion of cubic optical lattices, the potential near the trap center is approximately separable and well-described by

$$U(\mathbf{r}) \approx \sum_{q=x,y,z} v_q \sin^2 k_l q = V_x(x) + V_y(y) + V_z(z). \quad (2.64)$$

Each direction is characterized by the laser wave number $k_l = 2\pi/\lambda_l = \pi/a$, the laser wavelength λ_l , and the lattice separation distance a with minima at $q = 0, \pm a, \pm 2a, \dots$, and so on. In general all these quantities vary with the direction $q = x, y, z$ but we usually omit this additional notation for brevity. The trapping strengths are typically written $v_q = s \cdot E_r$ in units of the recoil energy $E_r = \frac{\hbar^2 k_l^2}{2m} = \frac{\hbar^2 (\pi/a)^2}{2m} = \frac{\hbar^2 \pi^2}{2ma^2}$. The approximately independent one-dimensional lattice potentials $V_q =$

¹⁷ Other platforms include, e.g., superconducting islands with electron Cooper pairs undergoing Josephson tunneling and helium atoms moving on a substrate [149].

$V_q(q)$ admit a separation of variables into $|\psi(\mathbf{r})\rangle = |\psi_x\rangle |\psi_y\rangle |\psi_z\rangle$ with single-particle Hamiltonians $\hat{H}_{0,q} = -\hbar^2 \partial_q^2 / (2m) + V_q$ and in the following we will focus on the x -direction and drop most of these subscripts. We shall find a suitable single-particle expansion for the Hamiltonian Eq. (2.36) that are in some sense site local. The single-particle energy eigenstates are useful in constructing these although they are themselves delocalized across the lattice.

Bloch's theorem for periodic crystalline structures, $V_x(x) = V_x(x + a)$, allows a decomposition of the $\hat{H}_{0,q}$ eigenstates (Bloch states) into

$$\epsilon_n^k(x) = e^{ikx} u_n^k(x), \quad u_n^k(x) = u_n^k(x + a). \quad (2.65)$$

That is, $u_n^k(x)$ has the same periodicity as $V_x(x)$ and is modulated by plane waves to form $\epsilon_n^k(x)$. The quantum number n is the band number and the wave number k is related to the quasi-momentum by $p = \hbar k$ which is unique up to a reciprocal lattice vector k_l . Although no new information is contained for k -intervals larger than k_l , it is customary to consider the first Brillouin zone

$$-k_l \leq k \leq k_l \quad \Rightarrow \quad -\frac{\pi}{a} \leq k \leq +\frac{\pi}{a}. \quad (2.66)$$

The width of the zone is obviously $2\pi/a = 2k_l$. The wave numbers k are not continuous due to the finite extent of the lattice (length $L = N_{\text{sites}}a$), giving a k -spacing of $\Delta k = 2\pi/L = 2\pi/(N_{\text{sites}}a)$. We can thus parametrize the valid values of k in the first Brillouin zone as

$$k_m = -\frac{\pi}{a} + m\Delta k = -\frac{\pi}{a} + m\frac{2\pi}{N_{\text{sites}}a} = \frac{\pi}{a} \left(2\frac{m}{N_{\text{sites}}} - 1 \right), \quad (2.67)$$

where $m = 0, 1, 2, \dots, N_{\text{sites}} - 1$ (we may possibly ignore an overcounting by 1 in the limit of large N_{sites}). From here we take $k \leftrightarrow k_m$ implicitly. Substituting in the Bloch states into the time independent Schrödinger equation and a few lines of calculation yields an equation for the periodic functions u_k^n

$$\begin{aligned} \left[\frac{\hbar^2}{2m} \left(-i\frac{\partial}{\partial x} + k \right)^2 + V \right] u_k^n &= \left[\frac{\hbar^2}{2m} \left(-\frac{\partial^2}{\partial x^2} - 2ik\frac{\partial}{\partial x} + k^2 \right) + V \right] u_n^k \\ &= E_n^k u_n^k, \end{aligned} \quad (2.68)$$

where the additional k terms come from the derivatives of e^{ikx} (both sides divide out e^{ikx} subsequently). Solving this equation for a fixed $k \in k_m$ is usually done by inserting the Fourier series expansion of u_k^n and V which yields a particularly small and simple system of equation for the few non-vanishing Fourier coefficients due to the simple $\sin^2$ periodicity [208]. This is repeated for all $k \in k_m$. In this way we may numerically obtain the energies E_n^k and the eigenenergy Bloch states just by multiplying with e^{ikx} . Example results of this procedure is shown in Fig. 2.5.

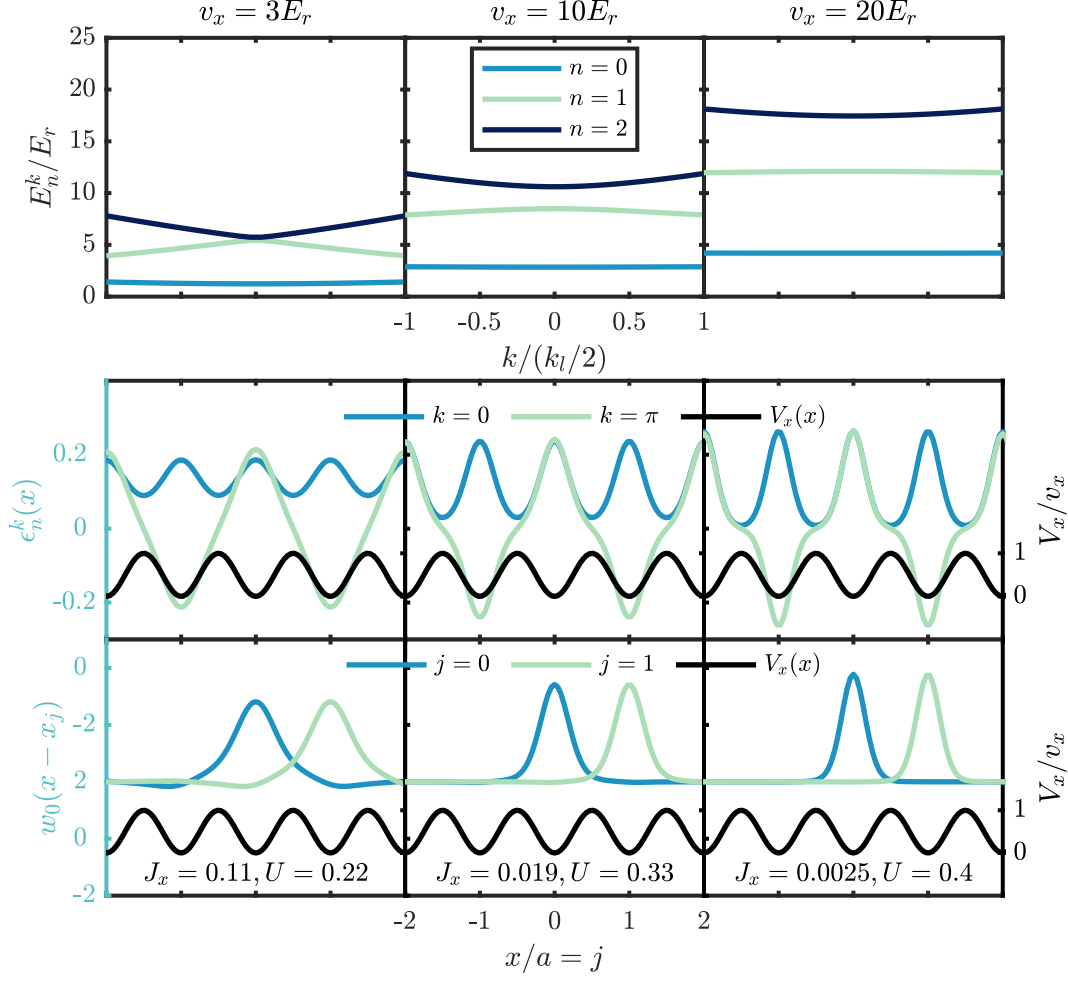


Figure 2.5: Single-particle states in the periodic lattice for three different depths v_x (column-wise). We take $v_y = v_z = 20E_r$ and $\lambda_l = 1064\text{ nm}$. Top panels: The band structure E_n^k for the lowest bands $n = 1, 2, 3$. Middle panels: Example Bloch states $\epsilon_n^k(x)$ for $k = 0, \pi$ (multicolored axis and lines) Bottom panels: Example Wannier states $w_0(x - x_j)$ for $j = 0, 1$ (multicolored axis and lines). The normalized potential $V_x(x)$ is shown with black lines and axes.

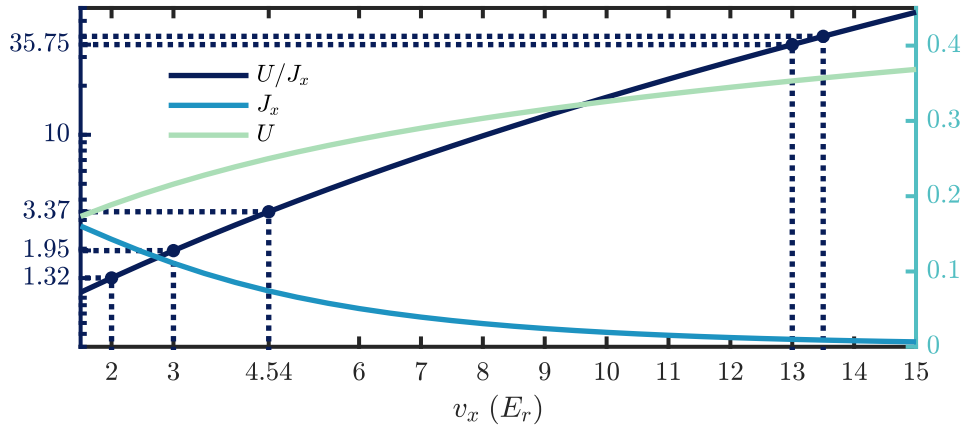


Figure 2.6: Adapted from Ref. [4]. The constitutive relations between v_x and the tunneling J_x and on-site interaction U strengths (multicolored lines and right axis). The ratio U/J_x is also shown (dark blue line and left axis) with a few U/J_x values are highlighted. Left to right they are as per [4]: the minimally allowed $v_x = 2E_r$, the v_x defining |SF>, the critical point of the one-dimensional |SF> $\rightarrow$ |Mott> phase transition, the v_x defining |Mott>, and the “maximally” allowed $v_x = 13.5E_r$ (corresponding to $U/J_x = 40.18$).

Note that the Bloch states are only defined up to a phase. We can use this gauge freedom to enforce the Bloch states to be real and positive at $x = 0$, i.e. requiring $\epsilon_n^k(0) = |\epsilon_n^k(0)|$. Suppose the phase was $e^{i\beta}$ instead — then multiplying with the phase $e^{i\alpha}$ we have

$$\epsilon_n^k(0) \leftarrow e^{i\alpha} \epsilon_n^k(0) = e^{i\alpha} |\epsilon_n^k(0)| e^{i\beta} = |\epsilon_n^k(0)| e^{i(\alpha+\beta)} \quad (2.69)$$

and we simply choose $\alpha = -\beta$ so that all the Bloch states constructively interfere near $x = 0$. The constructive interferences [209] are useful in defining the Wannier function defined on site j — i.e. centered and localized around the potential minimum at $x_j = j \cdot a$ — by summing all k contributions

$$w_n(x - x_j) = \frac{1}{\sqrt{\mathcal{N}}} \sum_{k \in k_m} e^{-ikx_j} \epsilon_n^k(x), \quad (2.70)$$

where $\mathcal{N}$ is a normalization constant. The energy obviously increases with n . Two neighboring Wannier states of the lowest band $n = 0$ are shown in Fig. 2.5 and are seen to be increasingly localized around site j and decaying exponentially away from the site center at a rate set by v_x . The Wannier functions are therefore good single particle states to expand on. Additionally the Wannier states are evidently connected by a simple spatial lattice site translation $\hat{t}(a)w_0(x - x_j) = w_0(x - x_{j+1})$ [209]. The set of all these Wannier functions, $\{w_n(x - x_j)\}$ for $n = 0, 1 \dots$ and $j = 0, \pm 1, \pm 2, \dots$, forms an orthonormal basis for the x -direction and we can similarly construct the Wannier basis in the y - and z -directions. These bases are joined to span the three-dimensional space with basis states on the form

$$W_{\mathbf{n}, \mathbf{j}} = W_{\mathbf{n}}(\mathbf{r} - \mathbf{R}_{\mathbf{j}}) = w_{n_x}(x - x_{j_x}) w_{n_y}(y - y_{j_y}) w_{n_z}(z - z_{j_z}), \quad (2.71)$$

where $\mathbf{n} = (n_x, n_y, n_z)$ and $\mathbf{j} = (j_x, j_y, j_z)$ are aggregated band and site indices for the different directions. This gives a suitable single-particle expansion for the field operator $\hat{a}_{\mathbf{r}} = \sum_{\mathbf{n}, \mathbf{j}} \hat{a}_{\mathbf{n}, \mathbf{j}} W_{\mathbf{n}, \mathbf{j}}$ for insertion in Eq. (2.36). At the same time we make the first of three simplifying assumptions, namely that we are in the ultracold limit and $v_x, v_y, v_z \geq v_{A_1}$ for some sufficient v_{A_1} such only the lowest bands are populated and transitions to higher bands are energetically suppressed (cf. Fig. 2.5). That is, we remain in $\mathbf{n} = \mathbf{0}$ and we henceforth drop that indexing with this understanding.

For the first (single-particle) term we obtain

$$\int \hat{a}_{\mathbf{r}}^\dagger \hat{H}_0 \hat{a}_{\mathbf{r}} d^3\mathbf{r} = \sum_{i, j} \left(\int W_i^* \hat{H}_0 W_j d^3\mathbf{r} \right) \hat{a}_i^\dagger \hat{a}_j \equiv - \sum_{i \neq j} J_{ij} \hat{a}_i^\dagger \hat{a}_j - \sum_{i=j} \mu_i \hat{n}_i. \quad (2.72)$$

Here, the matrix element μ_i is the chemical potential on site i while J_{ij} represent particle tunneling between sites i and j . Both are re-

solved further by distributing the Hamiltonian $H_0 = \sum_q H_{0,q}$ terms and factoring into independent integrals for each direction as

$$\mu_i = - \int W_i^* \hat{H}_0 W_i d^3 \mathbf{r} \equiv \mu_{x,i} + \mu_{y,i} + \mu_{z,i} \quad (2.73a)$$

$$\mu_{x,i} = - \langle w_{x,i_x} | \hat{H}_{0,x} | w_{x,i_x} \rangle, \quad (2.73b)$$

where we used $\langle w_{y,i_y} | w_{y,i_y} \rangle = \langle w_{z,i_z} | w_{z,i_z} \rangle = 1$ and

$$J_{ij} = - \int W_i^* \hat{H}_0 W_j d^3 \mathbf{r} \equiv J_{x,[ij]} + J_{y,[ij]} + J_{z,[ij]}, \quad (2.74a)$$

$$J_{x,[ij]} = - \langle w_{x,i_x} | \hat{H}_{0,x} | w_{x,j_x} \rangle \langle w_{y,i_y} | w_{y,j_y} \rangle \langle w_{z,i_z} | w_{z,j_z} \rangle, \quad (2.74b)$$

with similar definitions for $\mu_{y(z),i}$ and $J_{y(z),[i,j]}$.

For the second (interaction) term we obtain

$$\frac{g_{3D}}{2} \int \hat{a}_r^\dagger \hat{a}_r^\dagger \hat{a}_r \hat{a}_r d^3 \mathbf{r} \equiv \frac{1}{2} \sum_{ijkl} U_{ijkl} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l, \quad (2.75)$$

where U_{ijkl} represents the collisional interaction energy due to particles on sites i, j, k , and l . This matrix element can be factored into three independent integral contributions, one for each direction:

$$U_{ijkl} = g_{3D} \int W_i^* W_j^* W_k W_l d^3 \mathbf{r} \equiv U_{(ijkl)_x} U_{(ijkl)_y} U_{(ijkl)_z}, \quad (2.76a)$$

$$U_{(ijkl)_x} = \langle w_{i_x}, w_{j_x} | w_{k_x}, w_{l_x} \rangle. \quad (2.76b)$$

We now apply the second assumption, namely that the individual transverse trapping strengths fulfill $v_y, v_z \geq v_{A_2}$ (recall that the Wannier states are implicit functions of these) where v_{A_2} is some sufficiently high strength. This ordering means that tunneling dynamics conditioned on v_x are restricted to one-dimensional tubes, e.g. $\mathbf{i} = (i_x, 0, 0)$, since the Wannier functions belonging to different tubes have vanishing overlap and thus $J_{ij} = 0$ cf. Eq. (2.74). The only non-vanishing U_{ijkl} are likewise those in the same tube cf. Eq. (2.76). (The chemical potential is site-local and unaffected by this approximation.) Henceforth we shall therefore simplify the indexing as $\mathbf{i} = (i_x, 0, 0) \rightarrow i$. The non-vanishing matrix elements in the one-dimensional tube then read

$$J_{x,[ij]} = - \langle w_{x,i} | \hat{H}_{0,x} | w_{x,j} \rangle, \quad (2.77a)$$

$$U_{ijkl} = g_{3D} \prod_{q=x,y,z} \langle w_{q,i}, w_{q,j} | w_{q,k}, w_{q,l} \rangle. \quad (2.77b)$$

where we used $\langle w_{y,0} | w_{y,0} \rangle = \langle w_{z,0} | w_{z,0} \rangle = 1$ for the tunneling term. These matrix elements are translationally invariant for fixed index distances since the Wannier functions are related by lattice displacements in a and we may take the $i = 0$ site as a reference point. We now apply the third and last simplifying assumption, namely that $v_x \geq v_{A_3} > v_{A_1}$ such that the Wannier states are sufficiently localized

in the sense that each site is only appreciably tunnel-coupled with its nearest neighbors, $|i - j| = 1$, and all interaction energies except the on-site, $i = j = k = l$, are vanishing. This is known as the tight-binding limit, and the non-vanishing, isotropic elements read

$$J_x = -\langle w_{x,1} | \hat{H}_{0,x} | w_{x,0} \rangle, \quad (2.78a)$$

$$U = g_{3D} \prod_{q=x,y,z} \langle w_{q,0} | w_{q,0} | w_{q,0} \rangle. \quad (2.78b)$$

Values at different lattice depths v_x is seen in Fig. 2.5. The maximally allowed value shown there is imposed only by the control problem as described in Ref. [4]. Defining the operators $\hat{h}_{[i,i+1]}^J = -(\hat{a}_{i+1}^\dagger \hat{a}_i + \text{h.c.})$ and $\hat{h}_{[i]}^U = \hat{a}_i^\dagger \hat{a}_i^\dagger \hat{a}_i \hat{a}_i = \hat{n}_i(\hat{n}_i - 1)$ we finally obtain the result in Eq. (2.61a)

$$\hat{H}_{\text{BH}} = \hat{H}_x^J + \hat{H}^U = J_x \sum_{i=1}^{N_{\text{sites}}-1} \hat{h}_{[i,i+1]}^J + \frac{U}{2} \sum_{i=1}^{N_{\text{sites}}} \hat{h}_{[i]}^U, \quad (2.79)$$

for a finite number of sites N_{sites} . This is the simplest possible model for describing the boson lattice dynamics. The constant offset due to the chemical potential $\hat{H}^\mu = -\mu \sum_i \hat{n}_i$ been discarded as it is not relevant for the dynamics. This term is of course important in certain analytical studies, e.g. in establishing the phase diagram by meanfield theory [149]. We could also have included an additional, separable and site-dependent trapping potential, $\hat{H}_0 + \mathcal{V}(r)$, and this would yield an additional term on the form $\hat{H}^\mathcal{V} = \sum_i \mathcal{V}(x_i) \hat{n}_i$. The simplifying assumptions (only lowest band population, no tunnel coupling across one-dimensional tubes, and tight-binding limit) may also alleviated to include, e.g., more than just the lowest band [210], but these are not considered further here.

In passing, defining $s = v/E_r$ and taking for simplicity $v = v_x = v_y = v_z$ very deep, the integrals can be evaluated approximately as [86]

$$J \approx \frac{4}{\sqrt{\pi}} E_r s^{3/4} \exp[-2\sqrt{s}], \quad (2.80a)$$

$$U \approx \sqrt{\frac{8}{\pi}} k_l a E_r s^{3/4}, \quad (2.80b)$$

but these expressions are not preferred when the Wannier functions are available.

THE essence of quantum optimal control is to precisely steer physical processes governed by quantum dynamical equations of motion. The motivations discussed in Ch. 1–2 range from realizing novel quantum technologies to pursuing otherwise inaccessible basic research questions and anything in between. Similarly, the associated control problems span virtually both all imaginable physical platforms, be it atomic, molecular, photonic, solid solid-state, etc., and sizes of Hilbert space. It is quite notable how all of these very diverse applications fit into the same overall mathematical framework, despite that their individual treatment and appropriate angles of attack may in practice be vastly different.

The purpose of this chapter is to introduce the basic notions and methodologies within the quantum optimal control framework. In particular, the spirit of the presentation is to unfold a pragmatic and conceptually clear overview of central themes and ideas that constitute the foundations in the following manuscripts. Due to the format afforded by regular research articles the majority of these concepts are not — and cannot — be described in detail therein. Instead they are usually mentioned or introduced in a short paragraph, a sentence or two, or not at all. For some of these, e.g. concepts related to *optimal control strategies*, I have here attempted to provide collective mathematical definitions as a means to categorize and think about common patterns emerging from and pertaining to a variety of problems. I am unsure if these can in any way be construed as controversial — that is not their intent — and it may be possible to conjure up cases, pathological or not, where the given definitions are not quite appropriate.

This exposition is thus meant as an extensive topical review that, to the best of my ability and from my point of view, covers the fields’ building blocks and how they connect. As the chapter unfolds I will piece together the fairly decorated specification, *numerical iterative derivative-based open-loop quantum optimal control of closed systems*, given in the introduction. Each qualifier naturally implies at least one other alternative. I will provide a synopsis of these alternatives as appropriate and remark that, for each of these alternatives, a full account is worthy of a thesis in its own right.

The main workhorse in my studies is a derivative-based local optimization approach. Naturally this requires one to calculate certain derivatives and this is what leads to numerical working equations. In particular, it turns out that temporal discretization — which inevitably

will have to be addressed in a numerical treatment on a classical computer — has a significant impact on those calculations. However, the exact nature of these and similar considerations are not necessary for painting the big picture. Instead I have found it conducive to defer these very technical calculations and discussions to a separate chapter. This provides a much more focused narrative without being bogged down by intermittent derivations. Further to this point, most of the mathematical equations and expressions in this chapter are either definitions or attempting convey certain ideas.

The chapter is structured as follows.

Section 3.1 sets the stage by introducing the notion of controls that can steer the dynamical evolution and how one can ascribe quantifiable meaning to their effectiveness through task-encoding fidelity functionals. Optimal controls are extremizers of these functionals and we seek to find these with iterative optimization.

Section 3.2 introduces the remarkably convenient idea of control landscapes which is simply an illustrative abstraction of the encoding functional. Optimal controls correspond to peaks in the landscape. From this we can more straightforwardly define minimal process durations and optimal control strategies.

Section 3.3 describes — technical details notwithstanding — how the functional can be optimized, equivalently the control landscape maneuvered, using either local or global search methodologies.

Section 3.4 discusses seeding strategies, the mechanisms central to the efficiency of any iterative optimization algorithm. We define a general purpose parametric seeding structure with a clear exploration versus exploitation interpretation.

Section 3.5 expands on the parametric ansätze for seeding by establishing how the control itself can be fully parametrized and the emergent consequences thereof.

Section 3.6 lists a plethora of different techniques to accommodate any given experimental constraints (both in real and frequency space, etc.). To this end we re-frame the fidelity functional as an equivalent cost functional purely due to convention.

Section 3.7 presents further advanced developments of including robustness, homotopy methods, and imaginary physical boundaries.

Section 3.8 sheds light on several different perspectives to other types and aspects of quantum control, namely controllability, quantum speed limits, and coherent control. Finally, we discuss and argue about the merits and role of derivative-based local optimization and its results as a whole.

3.1 THE STATE TRANSFER PROBLEM

Any desired physical process can be encoded in an appropriate functional, F , which we may call the *fidelity* and typically $0 \leq F \leq 1$. The fidelity is a figure of merit (or objective) that quantifies success, how closely a given physically realized transformation resembles some desired target transformation. The goal is thus to maximize the fidelity, preferably as close to $F = 1$ as possible which signifies a perfect match between the realized and target transformations.

The actual structure and content of F depends both on the specific type of control task at hand and whether a state vector or density matrix formalism is considered for representing the quantum state. The work presented here has primarily focused on so-called pure state transfers (also called state-to-state transfers) under unitary Schrödinger(-like) dynamics. That is, the system is closed. In this setting, the goal is to facilitate the transfer $|\psi_{\text{ini}}\rangle \xrightarrow{T} |\psi_{\text{tgt}}\rangle$ where $|\psi_{\text{tgt}}\rangle$ is a specified target state, T is the process duration, and $|\psi_{\text{ini}}\rangle$ a specified initial state that evolves instantaneously according to

$$0 \leq t \leq T: \quad i|\dot{\psi}(t)\rangle = \hat{H}(t)|\psi(t)\rangle, \quad |\psi(t=0)\rangle = |\psi_{\text{ini}}\rangle, \quad (3.1a)$$

or equivalently in terms of the time-evolution operator (or propagator)

$$|\psi(T)\rangle = \hat{U}(T;0)|\psi_{\text{ini}}\rangle, \quad \hat{U}(T;0) = \mathcal{T} \exp\left(-i \int_0^T \hat{H}(t)dt\right), \quad (3.1b)$$

in units where $\hbar = 1$, $\dot{\psi} = \partial_t \psi = \partial \psi / \partial t$, and $\mathcal{T}$ denotes Dyson time ordering (equivalently Dyson series [181]). The system Hamiltonian is either linear or nonlinear corresponding to the usual Schrödinger or GP equation discussed in the previous chapter, respectively. Recall that the latter is a meanfield description of a BEC on the form $\hat{H}_{\text{gp}} = \hat{T} + \hat{V} + g|\psi|^2$ where g is the condensate self-interaction strength. The distinction and dimensionality is not important for the contents of this chapter.

At the heart of the theory is the notion of a *control*, call it $u(t)$, that somehow parametrizes the Hamiltonian,

$$\hat{H} = \hat{H}(t, u(t)) = \hat{H}^d(t) + \hat{H}^c(t, u(t)), \quad (3.2)$$

where we have separated into drift and control Hamiltonian contributions without loss of generality (the time-dependences will frequently be notionally suppressed as needed). One could make the case for a careful delineation, but we will indiscriminately¹ refer to $u(t)$ as a control parameter, a control function, a control field, a solution, a pulse, or simply a control. Thus, the control $u(t)$ is our manipulatory access point to influence the time translations generated by the Hamiltonian.

¹ “Kært barn har mange navne”

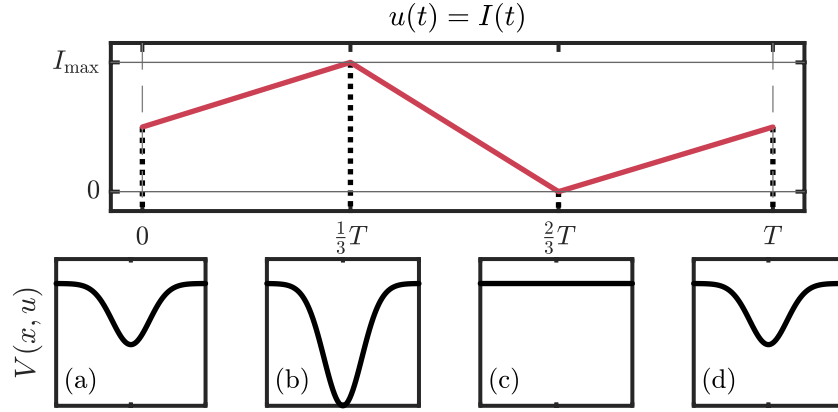


Figure 3.1: Adapted from the qualifying exam progress report. Example of control parametrized Hamiltonian $\hat{H}(u) = \hat{H}^d + \hat{H}^c(u) = \hat{T} + \hat{V}_{\text{tweezer}}(x, u)$. Here the control parameter is the laser intensity of an optical tweezer in Eq. (2.4) such that $V(x, u) \propto A \propto u$. Top: Example control function. Bottom: (a)-(d) V at the times marked in the top figure. Varying the control in time dynamically changes the potential in arbitrary units.

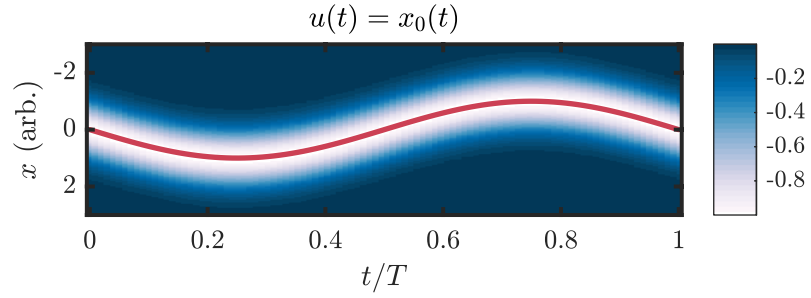


Figure 3.2: Same as Fig. 3.1, but with the control instead being the tweezer center position such that $V(x, u) \propto -\exp(-2(x - u)^2/w_0^2)$ where w_0 is a width. The red line is the example control function $u(t)$ and the vertical slices of color scale denotes the corresponding potential in arbitrary units.

It is of course entirely possible to entertain multiple controls and control Hamiltonians, $\{\hat{H}_q^c(\{u_k\}_{k=1}^K)\}_{q=1}^Q$, but in the expressions displayed here we consider $K = Q = 1$ to preserve clarity of the presentation and the extension $K, Q > 1$ is fairly straightforward and often obvious. The available controls and their interpretation varies from problem to problem. Usually, but not necessarily, a control is directly equal or proportional to a physical quantity such as laser intensity (Fig. 3.1), laser position (Fig. 3.2), or current in a wire. From a theoretical point of view there is no *a priori* mathematical distinction. It is, however, still valuable and usually even necessary to consider the control's relation to the corresponding physical quantity as evident from the examples. If a control u is directly equal to a laser intensity, $u(t) = I(t)$, we may need to enforce for any solution that $0 \leq u(t) \leq I_{\max}$ for some experimentally relevant $I_{\max}$ to obtain physical results. If our control is instead equal to the position of an optical tweezer, $u(t) = x_0(t)$, we are instead naturally limited by e.g. the size of the enclosing vacuum chamber. Quite generally we may therefore require appropriate boundedness $u_{\min} \leq u(t) \leq u_{\max}$. In a similar vein we may also want to limit the rate of change of the control to accommodate finite electronic response times, minimize the time-integrated laser power, or handle other such physical constraints. We will defer a detailed description of this to Sec. 3.6 and simply assume that these types of constraints can indeed be handled until then.

Very often, the control Hamiltonian is bilinear, $\hat{H}^c(u) \rightarrow u \cdot \hat{H}^c$, and in many theoretical expositions this is therefore assumed early on. We shall not do so here, because, if applicable, this limiting case can always be taken at the end. In my opinion, this is preferable since preserving generality often comes without mentionable additional effort and e.g. the situations in Figs. 3.1–3.2, which are bilinear and not bilinear, respectively, are treated on an equal footing.

Using Eqs. (3.1)–(3.2) we can concisely encode the state transfer problem in the following fidelity functional,

$$F[u(t); T] = |\langle \psi_{\text{tgt}} | \psi(T) \rangle|^2 \quad (3.3a)$$

$$= |\langle \psi_{\text{tgt}} | \hat{U}(T; 0) | \psi_{\text{ini}} \rangle|^2. \quad (3.3b)$$

Quite clearly, $0 \leq F \leq 1$, and $F = 1$ is obtained only if a particular control $u^*(t)$ correctly steers the evolution such that $|\psi_{\text{ini}}\rangle \rightarrow |\psi(T)\rangle = |\psi_{\text{tgt}}\rangle$ up to a global phase which is inconsequential for our purposes. Identifying such *optimal* controls, however, is usually a very challenging prospect. Invoking a momentary license for colloquialisms, solving² for “?” in the *forward problem*, $F[u(t); T] = ?$, is straightforward by Eq. (3.1) whereas in seeking to solve the *inverse*

² There is a certain allusion to the **P** versus **NP** problem in computational complexity theory [211]. If $\mathbf{P} \neq \mathbf{NP}$, it is comparatively easy to *verify* that a solution solves a given problem while it is possibly exceedingly difficult to *find* a solution that solves the problem.

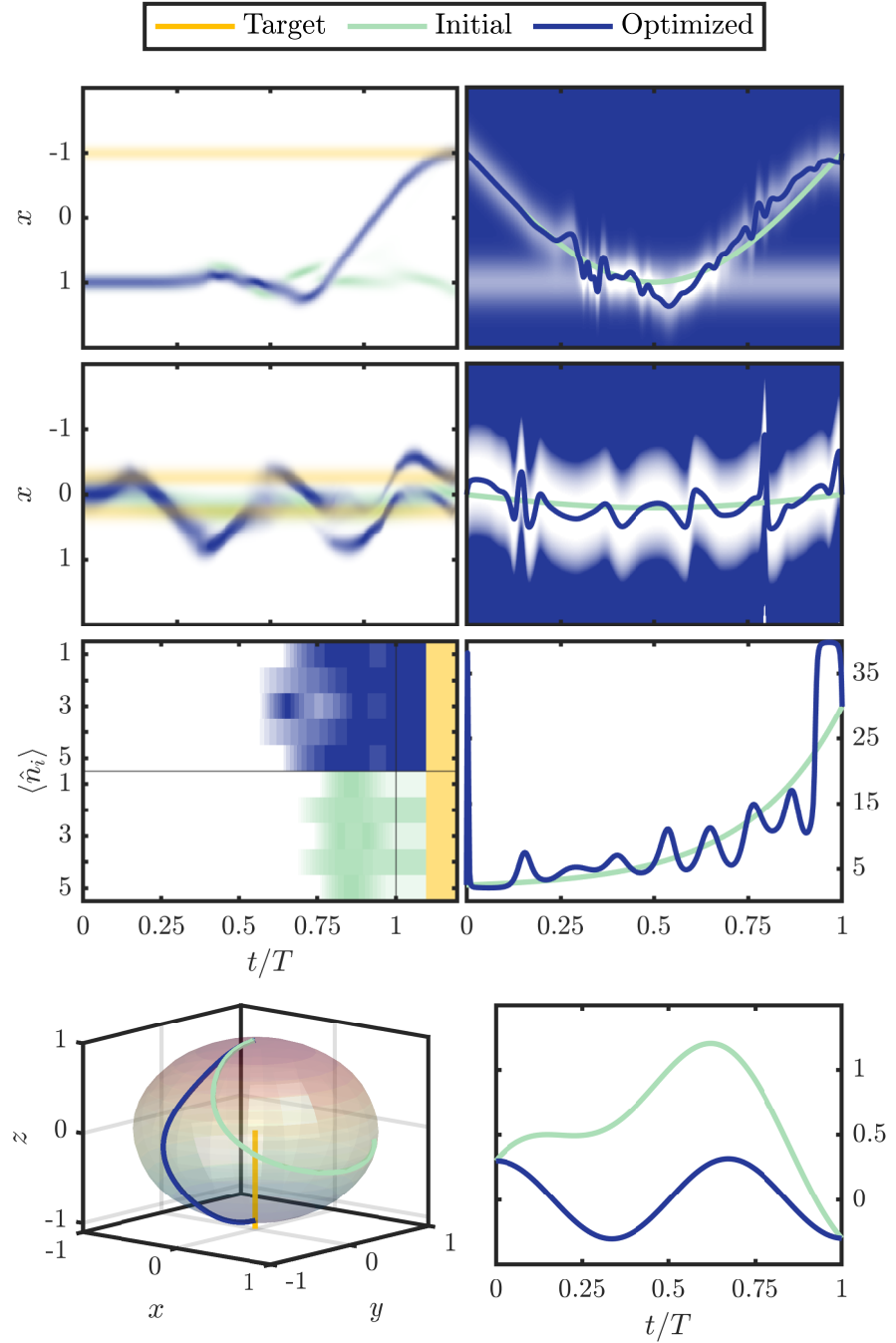


Figure 3.3: State transfer examples solved with the QENGINE [1]. Sub-optimal initial controls are optimized into optimal controls $F[u^*(t)] \gtrsim 0.999$ (right panels) and the target state and associated dynamics are visualized (left panels). **First row:** *Bring Home Water* [5]. A single particle resides in a static tweezer and is shuttled back to the initial position of the control tweezer Eqs. (2.58),(2.4) (left panel shows state densities and right panel colormap shows the combined potential with same y -axis). **Second row:** *Shake Up* [5]. A BEC is shaken from the ground state into the first excited state in an atom-chip potential Eqs. (2.60),(2.10). **Third row:** *Superfluid-Mott insulator phase transition* [4]. The two paradigmatic phases of the Bose-Hubbard model in an optical lattice is driven by controlling the tunneling and on-site interaction energy ratio Eqs. (2.61b),(2.6) (the colormaps are chosen to be sensitive around $\langle \hat{n}_i \rangle \approx 1$). **Fourth row:** *Spin flip* [3]. Transfer between $\hat{\sigma}_z$ eigenstates in the paradigmatic Landau-Zener model with a drifting X-drive and a controlled Z-drive (the left panel shows the Bloch sphere trajectories, $\mathbf{R}(t) = \sum_i \hat{r}_q \langle \sigma_q(t) \rangle$, $q = x, y, z$).

problem, $F[?; T] = 1$, we have up to this point no immediate tools. There may be any number of unique solutions, or what is worse, a solution may not even *exist* and instead we may have to settle for fidelity lower than 1. Among others, these issues relate to the process duration T entering as a parameter which one in turn is typically interested in minimizing for a multitude of reasons (beating decoherence times, increasing the number of quantum gate operations possible within some time frame, etc.). The ultimate speed with which we can traject in Hilbert space is intimately tied to Heisenberg's time-energy uncertainty relations as we discuss later in Sec. 3.8.

Nevertheless, we have reached the crux of the matter, namely the motivation and need for devising said optimization tools that are — somehow — capable of extracting optimal controls that realizes as close to $F = 1$ as possible. A variety of such state transfers optimized to this limit are shown in Fig. 3.3. Additionally, the prospective optimal controls should preferably drive the transfer near or at the so-called minimal duration $T_{\min}^{F=1}$ as defined in the section following,

$$F^* = \min_T \max_F F[u(t); T], \quad u^*(t) = \arg F^*. \quad (3.4)$$

The work presented in this thesis entails systematically mapping out the details of and asking questions related to Eqs. 3.4 for a variety of problems. A typical bulk result is shown in Fig. 3.4 where the infidelity $1 - F$ is mapped out as a function of T for many different optimization runs. It shows that there is a clear trade-off between duration and the maximally achievable fidelity. What is more, there appears to be a certain structure in the highlighted square where two individual exponential branches shoots off with different slopes and not much seems to be in between. How can we understand and interpret this behavior? And how do we obtain this type of result in the first place? Answering these questions and more is the program of the following sections in which we develop central concepts and local optimization tools. Before delving into this, a few remarks are in order.

The equivalent [181] Eqs. (3.3a)–(3.3b) are intentionally prescribed separate labels — this is because they invite very different approaches and the resulting working equations are, in fact, *not* equivalent when time discretization is inevitably imposed which will become evident in Ch. 4.

As opposed to the equations of motion which are dictated by physical laws, the functional encoding is purely a mathematical tool which is ours to freely tailor. For example, we could also have chosen e.g. $F = \Re \langle \psi_{\text{tgt}} | \psi(T) \rangle$ or similar³. The functional may even be expanded to include additional desired features rather than just the transfer fidelity, for instance the aforementioned control constraints. Other

³ With the caveat that certain choices may be subtly favorable, see e.g. Ref. [212].

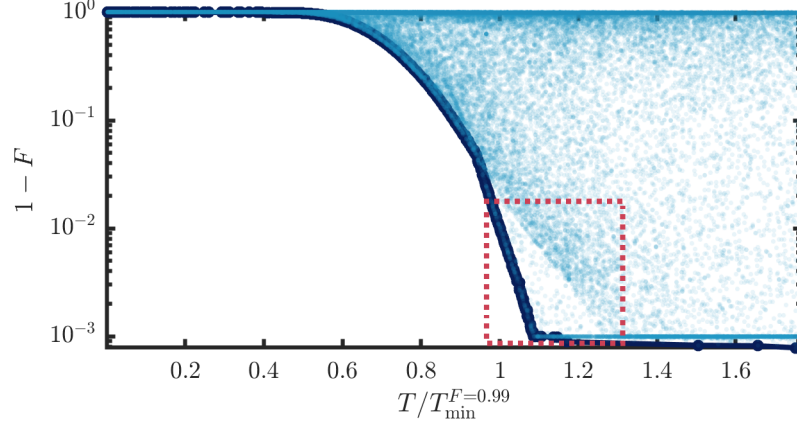


Figure 3.4: Adapted and simplified from [5]. Bulk results for the *Bring Home Water* state transfer problem described in Fig. 3.3. Light blue dots indicate $1 - F$ (lower is better) for each optimized solution and the dark blue dotted line shows the monotonically best results. Optimization was halted at $F \geq 0.999$. The red square highlights a curious branching of solutions.

types [213] of control tasks such as gate synthesis, trajectory optimization, and observable optimization, can be similarly encompassed by defining appropriate functionals. Relatedly, if one instead considers a density matrix formulation in Liouville space (allowing both pure and mixed states), an appropriate state transfer objective could be $F = \mathcal{N} \text{tr}(\rho_{\text{tgt}}^\dagger \rho(T))$ where $\mathcal{N}$ is normalization constant and tr is the trace operator. If the evolution is still unitary, the von Neumann equations of motion complementary to Eqs. 3.1 are $i\dot{\rho}(t) = [\hat{H}(t), \rho(t)]$ and $\rho(T) = \hat{\mathcal{U}}(T;0)\rho_{\text{ini}}\hat{\mathcal{U}}^\dagger(T;0)$. One could imagine, then, that the growing “zoo” of different cost functionals and physical state representations would lead to many separate algorithmic instruments. As far as local optimization is concerned, however, it turns out that these are trivially related as long as the evolution is unitary, i.e. $F = F[\hat{\mathcal{U}}(T;0)]$. In this sense, the focus on pure state transfer problems is therefore not really that restrictive. If the evolution is non-unitary and thus modeling a dissipative open system, Markovian or otherwise [101], the situation is somewhat more complex and will not be considered further here.

3.2 THE CONTROL LANDSCAPE

Although we are already in a position to describe the optimization of $F[u(t); T]$ we interject a few considerably simplifying ways of thinking about F itself and its properties.

As implied by its name, the mathematical characteristic [47, 214–216] of optimal controls $u^*(t)$ is that they are local optima of $F[u(t); T]$. Our goal is to maximize the fidelity and so local optima is in this context implicitly synonymous with local maxima — controls for which the first variation (related to the gradient) $D_{\delta u(t)} F$ vanishes and the second

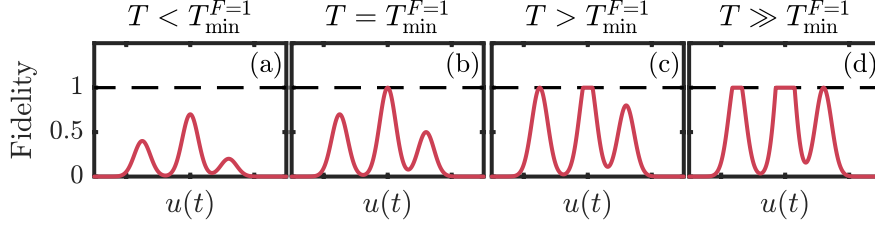


Figure 3.5: Adapted from the qualifying exam progress report. Abstract illustration of a fidelity control landscape at various durations T . The infinite number of solutions are abstractly collapsed onto the x -axis. Each point on this axis corresponds to a particular solution $u(t)$. The y -axis is the landscape height given by the mapping $F[u(t); T]$. (a): Below $T_{\min}^{F=1}$ no maxima are optimal (with respect to $F = 1$) and consists of three local traps. (b): Exactly at $T_{\min}^{F=1}$ (time)-optimal controls exists. In this example a single one. (c): Above $T_{\min}^{F=1}$ more optimal solutions emerge. (d) Far beyond $T_{\min}^{F=1}$ many optimal solutions exist and are contiguous to each other. In this example all maxima become optimal and no local traps remain.

variation (related to the Hessian) $D_{\delta u(t)}^2 F$ is negative (semi)definite such that [215]

$$F[u^*] \geq F[u^* + \delta u] \approx F[u^*] + D_{\delta u} F[\delta u] \Big|_{u^*} + \frac{1}{2} D_{\delta u}^2 F[\delta u] \Big|_{u^*}, \quad (3.5)$$

for any sufficiently small perturbation $\delta u(t)$. Typically, what we *really* are interested in are the *global* maxima, the best of the local maxima as measured by fidelity. These are oftentimes shortened to just *optimal controls* because “best optimal” projects a certain repetition. Further, *optimal controls* may also sometimes refer simply to a control that equals or exceeds a certain fidelity threshold which typically allows some slack to the actual global maximum value. These ambiguities in terminology are pragmatism that serves to allow much less cumbersome discussions and we will therefore also make liberal use of them in this thesis. Fortunately it is typically clear from the context what is meant and it helps that non-global local optima are called local “traps”.

One particularly convenient and visually appealing notion is that of the *control landscape* or *optimization landscape* [217]. It is allowed to us without any new information by imagining a landscape being spanned by all possible controls $u(t)$ with a height given by the corresponding functional value $0 \leq F[u(t); T] \leq 1$. Optimal controls then correspond exactly to the landscape’s peaks as shown in Fig. 3.5. This type of depiction and artistic license is useful because the overall insights garnered from both one-dimensional and two-dimensional renditions generalize well to the actual infinite dimensional space. As mentioned in the previous section, there may exist zero, one, or multiple optimal controls (with respect to $F = 1$) depending on the process duration T . While there are no guarantees, an overall trend is that most problems become “easy” in the sense of high-fidelity abundances as $T \rightarrow \infty$. This

is trivial if $|\psi_{\text{ini}}\rangle$ and $|\psi_{\text{tgt}}\rangle$ are adiabatically connected by sufficiently smooth ramping of $u(t)$, but nonetheless it still appears to nearly always hold true in practice outside of these special cases. A heuristic argument is that the more time available, the less “average progress per unit time” is necessary to reach $F = 1$ and the more Hilbert space trajectories are viable. At the same time, this implies that the control problem becomes increasingly difficult as $T \rightarrow 0$ and below some minimal duration, $T_{\text{min}}^{F=1}$, it is impossible to achieve unit fidelity (and similar for lower fidelity thresholds). Thus, T_{min}^F is the fundamental minimal process duration where at least one control is optimal with respect to the threshold F — such controls are also sometimes called *time-optimal* with respect to the given fidelity threshold. These different regimes are depicted in Fig. 3.5 for $F = 1$. For illustrative simplicity, here only the height but not the location of the individual peak centers changes. In practice both would generally be affected as it is the entire control landscape structure that transforms with T as we will directly observe in Sec. 3.5. These structures can also be tracked by a variety of useful metrics [218–220] that can be empirically estimated by probing the landscape and recording along local optimization trajectories. The metrics include landscape “twistedness”, probability of $F = 1$, distance between optimal solutions, and so on. Relatedly, certain metrics like these can be thought of as order parameters and tracking them one can then possibly ascribe different control landscape *phases* as a function of T [218, 221]. At the phase transitions the landscape may abruptly change characteristics.

3.2.1 Optimal Control Strategies

The questions invited by Eq. (3.4) concerns the identification of individual optimal controls. A naturally related, broader question is then how these relate to each other. For example, as T increases in Fig. 3.5 the middlemost peak attains $F = 1$ for a relatively wide range of closely lying controls and this bulk behavior is typically representative of some valuable physical insight to be gained. We can describe and group such contiguous local maxima as follows.

A reasonable conceptual requirement for the landscape abstraction is that it preserves distances metrics (e.g. Euclidean) such that the distance between two control functions $u'(t) - u(t) = \delta u(t)$ are also close on the abstractly collapsed axes with similar fidelities when $\delta u(t)$ is small. This implies that controls spanning e.g. a $F = 1$ “ridge/plateau” are in some sense “path-connected” to each other through a series of smooth, continuous deformations such that the fidelity remains unchanged along the path. In other words, if $u_{\text{ref}}(t)$ is a reference control on one such ridge, there exists a certain set of perturbations $\Delta = \{\delta u(t)\}$, and for each not-necessarily-small $\delta u \in \Delta$ there exists an associated continuous path function $P_{\delta u}(x)$ from the unit interval

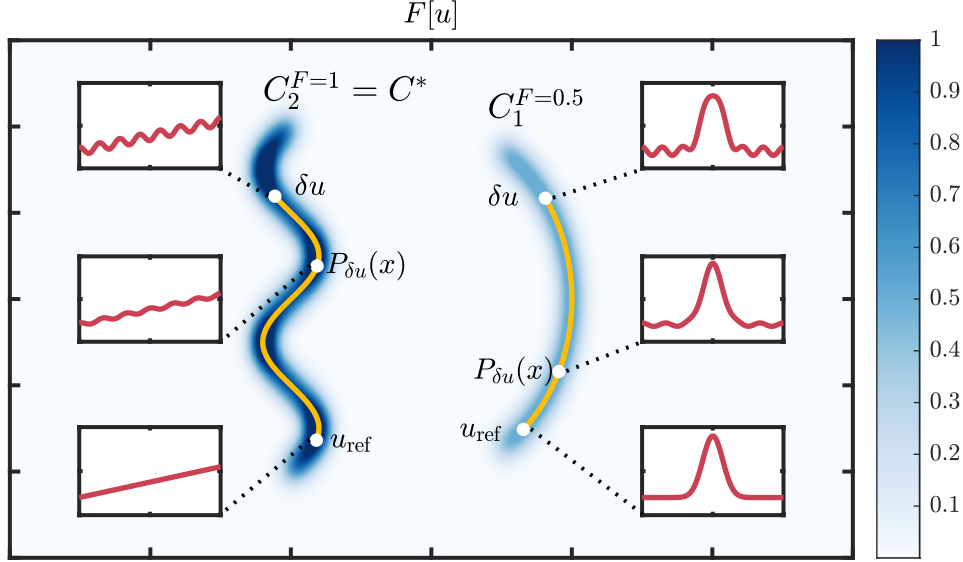


Figure 3.6: Abstract two dimensional illustration of optimal control strategies. Each point in the plane corresponds to a control $u(t)$ and the colormap shows the corresponding $F[u; T]$ at some fixed T . There are two distinct optimal control strategies $C_1^{F=0.5}$ and $C_2^{F=1}$ — isolated contiguous maxima ridges spanned by controls of similar shape. Any two points u_{ref} and δu hereon are smoothly path-connected (orange line) by a function $P_{\delta u}(x)$. The insets show the corresponding $u(t/T)$ (red lines, arbitrary units) for each of the highlighted example controls (white dots). The best strategy is $C^* = C_2^{F=1}$.

$[0, 1]$ to the control space with $P_{\delta u}(0) = u_{\text{ref}}$ and $P_{\delta u}(1) = u_{\text{ref}} + \delta u$ such that

$$C^{F=1}(T) = \{u = u_{\text{ref}} + \delta u \mid F[P_{\delta u}(x); T] = 1, \quad \forall x \in [0, 1] \}. \quad (3.6)$$

$\forall \delta u \in \Delta$

Any choice of u_{ref} on the same ridge leads to the same set of functions $C^{F=1}$. Note that this is essentially a definition for the local level set for contiguous optimal solutions. Now we can call $C^{F=1}(T)$ an *optimal control strategy* because it encapsulates a family of controls exhibiting similar characteristics⁴ that leads to unit fidelity. However, since unit fidelity may be fundamentally prohibited for a given T we straightforwardly generalize Eq. (3.6) with a variable fidelity condition, $C^F(T)$ (the set of controls are contiguous with locally optimal value F). Additionally, as in the example Fig. 3.5, there might be any number of optimal control strategies, $\{C_i^F(T)\}_{i=1}^n$, that are generally relevant at

⁴ At least locally, since the contiguity can extend arbitrarily far. The definitions do not explicitly assert that all members necessarily must be similar. However, if two parts of the same strategy are very dissimilar, potentially divisible into sub-strategies by other means, then that may speak to an even deeper meaning — the contiguity condition is strong and this in some sense then shows that two parts can smoothly be interpolated [81].

different durations. That is, we are likely mainly interested in the best one(s) at T ,

$$C^*(T) = \max_F \{Q = \{\{C_i^F(T)\}_{i=1}^n\}\}, \text{ such that } Q \neq \emptyset. \quad (3.7)$$

These ideas are abstractly illustrated in Fig. 3.6 where two optimal control strategies are spanned by quite different types of related controls⁵.

The control characteristics exhibited in any given strategy could, for example, be a certain number of bangs [4], a certain number of periodic oscillations [5], or the endpoint position of linear ramping segments [5]. Quite generally, regardless of what these characteristics might be, they are related to the relevant physical processes in a given problem. A reasonable conjecture is therefore that these characteristics are to some extent scale invariant, i.e. in some sense preserved or at most changes relatively slowly and continuously with $T \rightarrow T + \delta T$ such that the overall shape $u(t/T)$ of the function is preserved. For example, if $u_1(t) = t$ provides an adiabatic transfer for duration $T = T'$, so does⁶ $u_2(t) = t/2$ for duration $T = 2T'$. Despite being different functions they physically correspond to the same *strategy* but at different time scales. It is thus sensible to scale the controls as $u(t/T)$ to allow scale invariant comparisons in this type of analysis (in which case $u_1(t/T') = u_2(t/2T') = t/T'$ correctly coincides as the same strategy).

For a more concrete example of how these ideas can manifest and change with T we highlight the *Bring Home Water* problem from Figs. 3.3–3.4. The associated optimal control strategies are shown in Fig. 3.7. These were uncovered [5] by probing the landscape with a local optimization algorithm and analyses similar to the ones described below. The strategies encompass different physical interpretations and the best strategy $C^*(T)$ crosses over from one to the other at a certain duration T_c . Notably the best strategy in the high-fidelity regime, $C^*(T > T_c) = C_2(T)$, is apparently the lesser “visible” peak in the landscape as seen from the discovery multiplicities. The problem is treated in more detail in Ch. 8.

Equations (3.6)–(3.7) are descriptive but unfortunately not constructive. Practically uncovering the T -dependent hierarchy of strategies as in Fig. 3.7 varies quite prodigiously between problems and can be challenging. This is related in part to the discussion points and generalized definitions presented later when we consider seeding strategies and control parametrizations (Secs. 3.4–3.5). In any event,

⁵ The definitions could be extended similarly to before by relaxing the fidelity condition to just require being above some chosen threshold such that (1) each control strategy $C^{F>F'}$ captures all local contiguous controls exceeding F' (a set of level sets) and (2) all control strategies exceeding F' also captured by $C^{*,F>F'}$, respectively.

⁶ Assuming appropriate “non-destructive” drift Hamiltonian contributions, for example in the tweezer transport in Fig. 3.3.

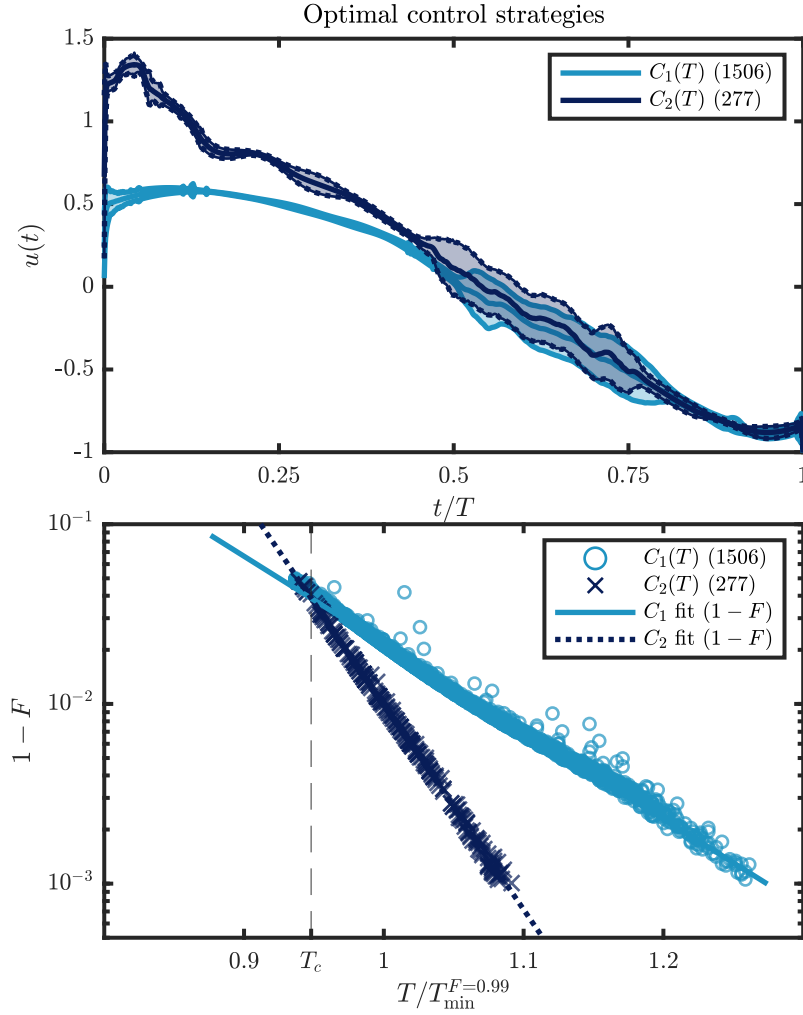


Figure 3.7: Adapted and simplified from [5]. Optimal control strategies in *Bring Home Water* (cf. Figs. 3.3–3.4). Top: mean and standard deviation of optimal controls found by an optimization algorithm from two distinct optimal control strategies $C_1^F(T)$ and $C_2^F(T)$ with a fidelity threshold of $F = 0.95$ (suppressed in the legend for readability). The parentheses show the respective discovery multiplicities. Bottom: $1 - F$ distribution (lower is better) of individual controls and fits. The best optimal control strategy is T -dependent with $C^*(T < T_c) = C_1$ and $C^*(T > T_c) = C_2$. Although the individually sampled controls vary in duration, their overall shapes $u(t/T)$ are very similar (very low standard deviation).

finding optimal control strategies is an iterative procedure. It often entails visual inspection of bulk optimization results either in time or frequency space — possibly sorted or transformed in some way — from which certain patterns may be apparent. If so, the next step is to capture the characteristics thereof such that individual controls can be correctly grouped⁷ into the corresponding strategies. Finally, inspecting the quantum dynamical evolution of states, observables, figures of merit, etc., from representative solutions can unveil a greater truth about the problem at hand. These insights can then be applied and tested in future optimizations until satisfactory. The aforementioned landscape metrics might also be useful in this endeavor. Hence, albeit inherently more involved, this type of meta-analysis is very valuable because it enables a more complete problem characterization which cannot be concluded on the basis of individual optimal solutions in a vacuum.

3.3 TRAVERSING THE LANDSCAPE BY ITERATIVE OPTIMIZATION

The auspicious depiction of control landscapes is a useful guiding aid, but it is nevertheless an abstraction unless the control is itself parametrized as described later. Identifying the optimal controls and strategies by mapping out the control landscape for even a single T with exhaustive search is impossible as there are infinitely many points $u(t)$ to try. This is also practically true in a numerical setting where the controls are discretized in time. Additionally, as mentioned previously, analytical solutions to these inverse problems are only possible in very few situations [222]. Instead we may turn to iterative optimization algorithms to locate these maxima and associated optimal controls and possibly strategies. Broadly speaking these can be divided into two classes: local optimizers and global optimizers. Although the remainder of the thesis concerns the former, we briefly review both without delving far into technical details.

3.3.1 Local optimizers

The unifying trait of local optimizers is that they greedily improve an initial solution $u^{(0)}(t)$ which we will also call a *seed* by producing a sequence of points with increasing quality,

$$u^{(0)} \rightarrow u^{(1)} \rightarrow \dots \rightarrow u^{(k)}, \quad (3.8a)$$

$$F[u^{(0)}] \leq F[u^{(1)}] \leq \dots \leq F[u^{(k)}]. \quad (3.8b)$$

⁷ The grouping can, e.g., be attempted by cluster analysis [5] or similar. Most clustering algorithms attempt to group elements based on a similarity condition. In our case the condition would jointly be contiguity in some (see Sec. 3.5) control landscape and fidelity as per our optimal control strategy definitions.

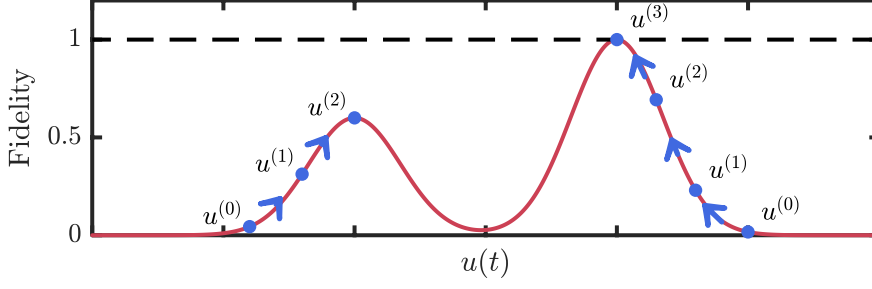


Figure 3.8: Adapted from the qualifying exam progress report. Multistarting a local optimizer from different seeds $u^{(0)}$ in the control landscape typically leads to different optima.

In particular, we will consider *derivative-based* local optimizers. Among these are the [GRAPE](#) [59], [GROUP](#) [165], [GOAT](#) [223], and [KROTOV](#) [224] algorithms (discussed in the next chapter). An evocative picture is, then, that we are climbing the control landscape as illustrated in Fig. 3.8 by taking small steps according to the update rule

$$u^{(k+1)} = u^{(k)} + \alpha^{(k)} p^{(k)}. \quad (3.9)$$

where $\alpha^{(k)}$ is a suitable step size along the search direction $p^{(k)}(t)$ for the k 'th iteration. The search direction is based on local derivatives typically to first order (gradient) and second order (Hessian). This is a standard methodology in optimization literature [214] and there several different ways of obtaining these derivatives. The success criteria [3] for any of these is that they must be (1) precise and preferably exact with respect to the “true” derivatives, and (2) efficient to calculate. This thesis takes a model-based stance, specifically by deriving analytically closed forms that can be implemented numerically with very high efficiency. Another appealing model-based numerical approach is that of automatic differentiation [225, 226] which is attributable to the differential decomposition of any computational program through the chain-rule. This approach is very flexible, but is generally comparatively slower on runtime than an analytically closed form. Yet a third possibility is that of finite differences. This is the most simple yet crude approach. It can either be obtained by model-based means or by model-free means directly from experiment. The versatility on offer, however, typically does not outweigh the associated drawbacks such as instability and inherent inexactness [70, 214] — especially, experimental estimation of just a single gradient is often just too expensive and/or impractical. Nevertheless, finite difference derivatives are still very useful in the numerical setting as a means to verify analytical derivatives.

Assuming the derivatives are sufficiently exact, the exploitative nature of derivative-based algorithms makes them exceptionally efficient

in locating the local maxima: in principle⁸, any seed is deterministically mapped to the nearest maximum wherefore this is also called an *attractor*. The complication to this local-search paradigm is obviously that they are severely limited in exploring the landscape. The exploitation versus exploration trade-off for non-convex problems is a well established notion in optimization literature [214, 227]. At least some amount of exploration is necessary since the local maxima found by the local optimizers can be (and often is) local traps⁹ cf. Fig. 3.5 (a)–(c). A conceptually simple way of introducing exploration is through multistarting [233], i.e. starting a local optimizer at multiple different seeds $u^{(0)}$ in the optimization space as illustrated in Fig. 3.8. In this way we are able to probe different parts of the landscape. This locates a number of possibly different local maxima and the best one(s) of these can then be taken as an approximation to the global maximum. For instance, only if $F = 1$ is this guaranteed to be a true global maximum. Conversely finding the best solution(s) maxima to have $F < 1$ does not preclude the existence of a $F = 1$ global maximum as seen in Fig. 3.5 (b)–(c) and similar for other fidelity thresholds. As a result of this, any optimization methodology, derivative-based local optimizers or otherwise, is only able to estimate an upper bound of $T_{\min}^F$ in the general case.

3.3.2 Global optimizers

Another class of optimizers are so-called global optimizers. As implied by the name these attempt to locate the overall global maxima rather than the initially nearest local maximum. This is naturally an enticing intention, but as with the local-search paradigm, success is not a given.

As opposed to local optimizers, which typically maintains and sequentially improves a single solution as in Eq. 3.8, global optimizers generally maintain a set of candidate solutions (a population) which is then updated according to a set of rules. To this end there are many algorithmic families and even more variants within these. Perhaps the most notable examples include the genetic evolutionary-type [234] (e.g. CMA-ES and differential evolution), simulated annealing [235], and swarm-based optimization [236] (e.g. particle swarm and ant colony).

The exact details of these are very diverse, but they all, roughly, share at least one or more of the following overall characteristics.

⁸ Numerical and implementation details such as the inexact line search discussed later introduces the possibility of “skipping over” the true nearest attractor.

⁹ Starting with Ref. [228], it has been posited that all control landscapes are, under certain conditions, always either trap free or alternatively zero everywhere. Being trap free is the same as being convex — meaning all local minima are also global and with $F = 1$ as e.g. in panel (d) of Fig. 3.5. Analytically based arguments for and against (e.g. in Refs. [229, 230] and Refs. [231, 232], respectively) have been made. In my numerical studies, local traps are indeed a very common occurrence and the necessary conditions to fulfill the requirements of Ref. [228] are perhaps too ideal.

Initially, a wide net of candidate solutions is cast through some form of random sampling and are evaluated based on some figure of merit such as e.g. fidelity. Note that this type of approach therefore does not eliminate the need for multistarting. The population is then updated by culling, adding, modifying, or combining individual candidates, usually with a certain degree of stochasticity. Then the population is evaluated anew, and this process continues. A typical feature is that the stochasticity decreases with the number of iterations, thus naturally converging on more and more narrow regions of the input space until some specified criteria has been met. In the narrow convergence stage the algorithms effectively become local optimizers but without the theoretical guarantees granted by true local search methodologies. A proposal for embracing this is to combine the two approaches [5, 163] — this is sometimes also referred to as a *memetic* extension.

Global optimizers are also direct search methods. That is, they rely only on evaluating the objective function(al) itself and not any of its derivatives. The most famous direct search method is the simplex Nelder-Mead [237]. Direct search methods are nearly always driven by heuristics which hampers convergence rates. They may nevertheless be advantageous in noisy environments as experimental setups tend to be. Another advantage is that the rules are relatively intuitive and easily implemented compared to for example model-based derivatives. As such, they work fairly well [238] off-the-shelf and can also encompass the closed-loop experimental setting more naturally [239].

3.4 SEEDING STRATEGIES

Locating the nearest local maximum is the trait of local optimizers. The success of this paradigm in particular is therefore inherently tied to the chosen *seeding strategy* responsible for generating new seeds. For example, the landscapes illustrated in Figs. 3.5 and 3.8 are fairly benign: if the seed generation is uniform in the control landscape of Fig. 3.8 there is about fifty-fifty chance that any generated seed would optimize to the global maximum or to the local trap. The landscape is, however, seldomly that generous in practice and can be very rugged with many local traps — that is why there are many dots separated from the dark blue line of best results in Fig. 3.4. As argued earlier, this typically becomes especially true near $T_{\min}^{F=1}$ and a uniform sampling, although by definition all-encompassing, is therefore generally not particularly effective as illustrated in Fig. 3.9. Instead it would be preferable to devise a seeding strategy that covers only the “good” regions of control space.

At first glance we may broadly characterize a “good” region as one where there are on average many “good” seeds $u^{(0)}$ that upon optimization become high-fidelity solutions. We can nevertheless quite effortlessly provide a more refined definition based on the optimal

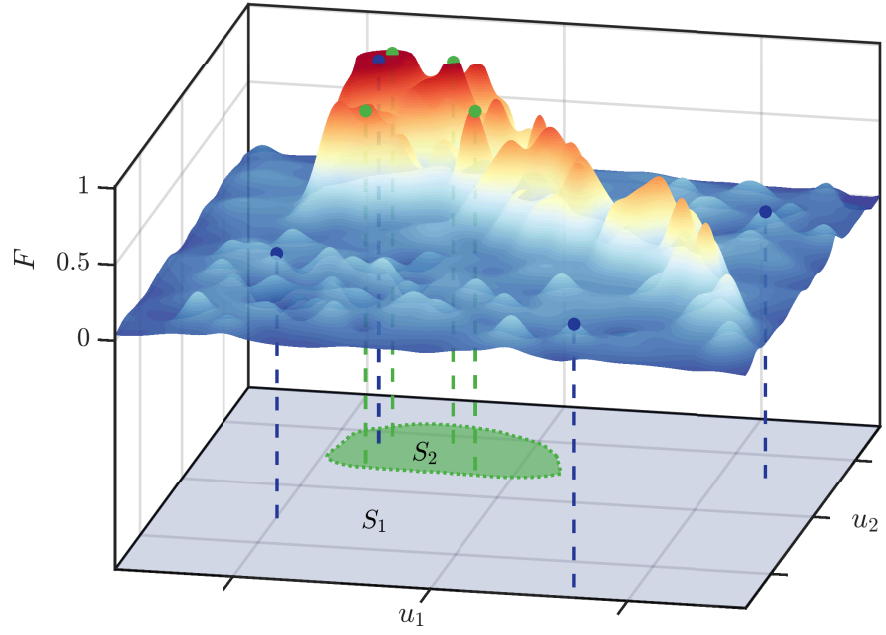


Figure 3.9: Adapted from [5]. A control landscape with two controls, $\{u_1(t), u_2(t)\}$, each visualized as spanning an axis. Only a small region of the control space is optimal (with respect to $F \approx 1$ in $C^*(T)$) and otherwise contains many local traps — the relative efficiency of a local hillclimbing optimizer is thus strongly correlated to where it is seeded. The dots denote controls that have been optimized to their local maximum and the color of a dot indicates which seeding strategy produced it. Purple full region: Uniform random seeding strategy S_1 . Obtaining $F \approx 1$ is possible, but with low probability. Green smaller region: Specialized seeding strategy S_2 . Obtaining $F \approx 1$ is possible with high probability, but not guaranteed.

control strategies discussed in Sec. 3.2.1. In this regard a good seeding strategy S is one that generates $u^{(0)}$ with comparatively high probability such that local optimization produces elements of the best control strategy(ies), $u^{(0)} \rightarrow \dots \rightarrow u^{(k)} \in C^*(T)$. The best possible seeding strategy is of course $S = C^*(T)$ such that no optimization is required ($k = 0$). This is putting the horse before the cart, however, because $C^*(T)$ is what we are trying to actually learn through our optimization efforts and we therefore initially instead have to try and guess what a good strategy could be. To this end we can be inspired to propose a general parametric seeding ansatz,

$$u(t; \theta) = u_{\text{ref}}(t) + \delta u(t; \theta) = u_{\text{ref}}(t) + \sum_{n=1}^N f_n(t; \theta), \quad (3.10)$$

where $u_{\text{ref}}(t)$ is a reference function, $\delta u(t; \theta)$ is a perturbation assembled from a set of parametrized functions $f_n(t; \theta)$, and $\theta = [\theta_1, \dots, \theta_M]$ is a set of values individually generated from some random distribu-

tion $\theta_m \sim p_m$ bounded by $\theta_{m,\min} \leq \theta_m \leq \theta_{m,\max}$. A seeding strategy can then be defined as the set of all possible generatable functions,

$$S = \{u(t; \theta) = u_{\text{ref}}(t) + \delta u(t; \theta) \mid \theta_{\min} \leq \theta \leq \theta_{\max}\}, \quad (3.11)$$

and S^* as the set of functions attained by mapping each element of S to its nearest attractor. Note that generally $S \neq S^*$ since the optimizer is free to “escape” the parametrization. Later in Sec. 3.5 we discuss the possibility of guaranteeing $S = S^*$ i.e. that the control stays within the chosen parametrization upon optimization. This can be both advantageous or disadvantageous depending on the context as is likewise elucidated.

The definition (3.11) highlights the exploration versus exploitation trade-off. The interpretation is that $u_{\text{ref}}(t)$ anchors a particular point in the control landscape and $\delta u(t; \theta)$ then explores the vicinity (the hypervolume spanned by $S \cup S^*$) as exemplified by the colored control space regions of Fig. 3.9. The reference control could be inspired adiabatically, analytically, empirically or by whatever other means sensible to a domain-expert within the context of a given problem. The exploration extent is set entirely by the parametrization. A few standard seeding possibilities, illustrated in Fig. 3.10 and displaying quite different characteristics, are as follows:

Piecewise constant — Define the indicator function I and divide the time interval $[0, T]$ into M regularly δt -spaced sub-intervals r_m ,

$$I(b) = \begin{cases} 1, & \text{if boolean predicate } b \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (3.12)$$

$$r_m = [t_m, t_m + \delta t], \quad t_m = (m - 1)\delta t, \quad 1 \leq m \leq M. \quad (3.13)$$

Then we can write piecewise constant perturbations with $N = M$ on the form

$$f_n^{\text{pwc}}(t; \theta_n) = I(t \in r_n) \cdot (\kappa \theta_n), \quad u_{\min} \leq \theta_n \leq u_{\max}, \quad (3.14)$$

where $\in$ is understood as a boolean membership test, $\theta = [\theta_1, \dots, \theta_M]$ and typically $\theta_n \sim \text{uniform}(u_{\min}, u_{\max})$. Note that we could also principally take r_m as a parameter and thus vary the interval lengths into a non-regular grid. The amount of exploration away from $u_{\text{ref}}(t)$ is set by $0 \leq \kappa \leq 1$ where $\kappa = 0$ is no exploration and $\kappa = 1$ is most exploration. For the latter and when $u_{\text{ref}} = 0$ we obtain a uniform random distribution in all M intervals $f_n^{\text{pwc}} \rightarrow f_n^{\text{uniform}}$.

Bang-bang — Bang-bang structures are regions of alternating extremal values of the control,

$$f_n^{\text{bang}}(t; \theta) = I(t \in b_n) \cdot h_n, \quad h_n \in \{u_{\min}, u_{\max}\}, \quad (3.15)$$

where $b_n = [c_n - \frac{w_n}{2}, c_n + \frac{w_n}{2}]$ and $\theta = [c_1, \dots, w_1, \dots, h_1, \dots]$ are the individual centers, widths, and heights of the bangs.

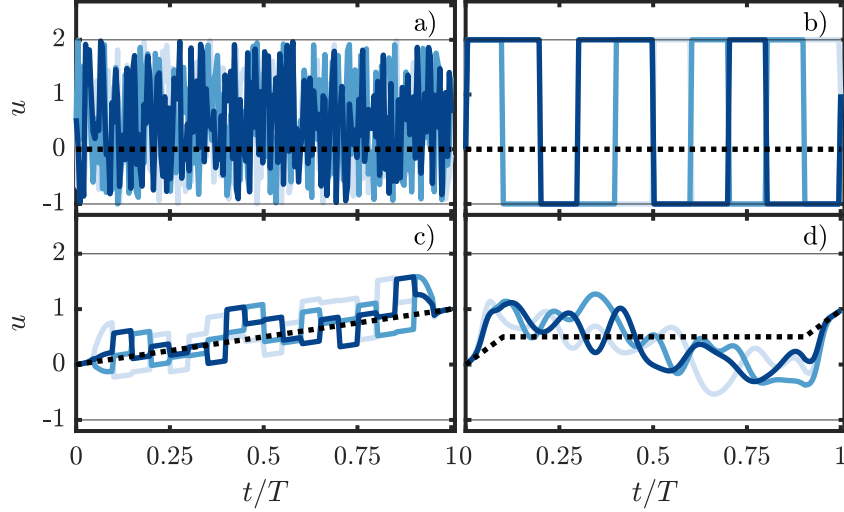


Figure 3.10: Seeding strategy examples with parametric functions (3.14)–(3.16). The solid lines are individual seeds, the black dotted line is the reference control, and the solid horizontal lines denote the control boundaries. a) Uniform random, f_n^{uniform} . b) Bang-bang, f_n^{bang} . c) Piecewise constant, f_n^{pwc} . d) Fourier, f_n^{Fourier} .

Fourier — An expansion on Fourier-like components with

$$f_n^{\text{Fourier}}(t; \theta) = c_n^c \cos(\omega_n^c t + \phi_n^c) + c_n^s \sin(\omega_n^s t + \phi_n^s) \quad (3.16)$$

where $\theta = [c_1^c, \dots, \omega_1^c, \dots, \phi_1^c, \dots, c_1^s, \dots, \omega_1^s, \dots, \phi_1^s, \dots]$ are expansion coefficients, frequencies, and phase shifts for the individual oscillatory components. Usually the frequencies are perturbed harmonics, e.g. $\omega_n^c = 2\pi((n-1) + \chi_n^c)/T$ where $|\chi_n^c| \leq 1/2$, and the expansion coefficients attenuated as e.g. $c_n^s \propto \exp(-n)$.

This list of examples is of course not exhaustive. For instance, we could define piecewise linear segments similarly to the piecewise constant ones. Additionally, any of these may also be subject to further modifications such as clamping values to inside the control boundaries, introducing an overall shape function $\delta u(t) \rightarrow \Gamma(t)\delta u(t)$ such that the perturbation goes to zero in the limit $t \rightarrow 0, T$ (as in panels c)–d)), or imposing an overall scaling of the perturbation or full seed. One can then also add heuristics on a higher layer, such as *pre-selection*: this consists in generating a number of seeds, calculating their individual fidelities, greedily selecting the seed(s) with highest fidelity for optimization, and discarding the others. This adds a safeguard against spending optimization time on seeds that initially looks unpromising, but it is obviously fairly naïve and may work well for some strategies and not for others [5].

As far as obtaining the fundamentally best possible results is concerned, devising a good seeding strategy is often as much an art as it is a science and is very often challenging — it is both strongly

problem dependent and it can be tricky and/or hard to programmatically implement even relatively simple conceptual ideas without any guarantee of significantly or even notably improved results. It is often a good idea to try out different seeding strategies and analyze the results as described in the optimal control strategy discussion following Eq. 3.7. Patterns emerging from such analyses can inspire new parametrizations for a seeding strategy, introducing a layer of meta-iteration that hones in on what is hopefully $C^*(T)$.

3.5 CONTROL PARAMETRIZATION

Employing parametric ansätze for seeding strategies are valuable as it allows us to probe certain specified regions in the control landscape. Perhaps it is empirically indicated or better yet known that one or more optimal control strategies are captured by one or more of these parametrizations or perhaps there is an experimental constraint that disqualifies certain control features which can be thusly circumvented. For the same reasons it may be valuable to *stay* within the hypothetical appropriate parametrization. However, the update rule $u^{(k+1)} = u^{(k)} + \alpha^{(k)} p^{(k)}$ Eq. (3.8) is contrary to this wish because the step $\alpha^{(k)} p^{(k)}(t)$ is constructed directly from derivatives on full control $u(t)$ and not on the parametrization parameters θ for the seed $u^{(0)} = u(t; \theta)$. Taking any step in this way is virtually guaranteed to exit the parametrization. Conceptually this not difficult to rectify as we can simply define an analogous update rule,

$$\theta^{(k+1)} = \theta^{(k)} + \alpha^{(k)} p^{(k)}, \quad (3.17)$$

where the search direction $p^{(k)}$ is now constructed from derivatives with respect to the parameters θ . These derivatives are connected to those of Eq. (3.8) by the chain-rule.

This type of update elicits a sequence of the form

$$\theta^{(0)} \rightarrow \theta^{(1)} \rightarrow \dots \rightarrow \theta^{(k)}, \quad (3.18a)$$

$$u(t; \theta^{(0)}) \rightarrow u(t; \theta^{(1)}) \rightarrow \dots \rightarrow u(t; \theta^{(k)}), \quad (3.18b)$$

$$F[u(t; \theta^{(0)})] \leq F[u(t; \theta^{(1)})] \leq \dots \leq F[u(t; \theta^{(k)})], \quad (3.18c)$$

which by construction remains within the chosen parametrization. This has a few related consequences. Firstly, we can now consider θ as *the* control, a finite M -dimensional vector — as opposed to a function in the continuous variable t — that can take any value within the specified control limits. Although elements of θ are also principally continuous variables, this still represents a huge dimensionality reduction in most cases because M is typically on the order of tens or hundreds. In fact, we obtain an *entirely new* control landscape,

$$F[u(t)] \rightarrow F[u(t; \theta)] \equiv F(\theta), \quad (3.19)$$

which is a regular *function* (i.e. not a *functional*) in M variables. Correspondingly the parameters can be considered as M spanning axes for the landscape.

Why is it emphasized that $F(\theta)$ is a *new* landscape? It is certainly just a dimensionality-reduced “subspace” of the full landscape $F[u]$. While it is true that the parametrization cannot change *heights* (value mappings) it can, however, shift the relative *positions* and thus overall topography. Points that are distant to each other in the $F[u]$ landscape may therefore be close in the $F(\theta)$ landscape and vice versa. This is cause for reflection in our definition of optimal control strategies. In Sec. 3.2.1 they were defined as contiguous optima in $F[u]$, but it is perhaps more correct to define an optimal control strategy as one that has contiguous optima in *some appropriate* parametrization $F(\theta)$. This is just a generalized definition as taking e.g. $\lim_{\delta t \rightarrow 0} F(\theta) = F[\theta(t)]$ for f_n^{uniform} and similar for other complete bases recovers the continuous description amounting to a re-labeling of the control variable from $u \rightarrow \theta$.

The generalized definition bears with it double edged ramifications. On one hand it may not be obvious what the appropriate parametrization is. On the other hand — if it can indeed be identified — it also means that the control problem is “easy” when viewed in the appropriate parametrization. In that case, physical insights can be gleaned from the parametrization and the associated dynamics, thus giving a more complete characterization of the problem.

3.5.1 An instructive toy problem

Let us consider the paradigmatic [240] Landau-Zener Hamiltonian as in Fig. 3.3 with a controllable Z-drive, $\hat{H} = \hat{\sigma}_x/2 + u(t) \cdot \hat{\sigma}_z/2$, and the state transfer target $|\psi_{\text{ini}}\rangle = |\uparrow\rangle \rightarrow |\psi_{\text{tgt}}\rangle = |\downarrow\rangle$ (eigenstates of $\hat{\sigma}_z$). Let us parametrize the control as a piecewise constant function with either $M = 1$ or $M = 2$ parameters cf. Sec. 3.4, i.e.

$$u = u(t; \theta_1) = \theta_1, \quad \text{and} \quad u = u(t; \theta_1, \theta_2) = \begin{cases} \theta_1, & \text{if } t < T/2, \\ \theta_2, & \text{otherwise.} \end{cases} \quad (3.20)$$

with a vanishing reference $u_{\text{ref}} = 0$.

The control landscapes [218, 222] $F(\theta_1; T)$ and $F(\theta_1, \theta_2; T)$ can then be charted and mapped exactly without being an abstraction as shown in Figs. 3.11–3.12. They capture many of the ideas up to now in action including that the control landscape transforms with T . Incidentally, this is one of the few cases where time-optimal solutions are analytically available [222], here at $T_{\text{min}}^{F=1} = \pi$ with $u^*(t; \theta_1 = \theta_2 = 0) = 0$. In fact this control is optimal periodically at durations $T^{F=1} = (1 + 2n)\pi$ for $n = 0, 1, \dots$ as the state makes a full revolutions on the Bloch sphere (Fig. 3.3). This is clearly seen in the $M = 1$ case. In between

these maxima, however, there are large intervals without $F = 1$ optimal solutions. This is not the case for $M = 2$ where $F = 1$ optimal solutions are indeed available for all shown $T \geq T_{\min}$, but there are also substantially more local traps as $T \rightarrow \infty$. It is emphasized that these are mainly due to the quite restrictive parametrization — the errors made can be trivially adjusted for by increasing M (similar to when $M = 1 \rightarrow 2$). The takeaway is that the problem can be easy when viewed in the right parametrization/control landscape, but the exact form hereof may be challenging to ascertain.

3.5.2 Chopped Random Basis (CRAB)

Certain parametrizations on the proposed form $u(t; \theta) = u_{\text{ref}} + \sum_n f_n(\theta)$ is also often called a CRAB [241]. The name strictly applies for popular choices like $\{f_n^{\text{Fourier}}(t; \{c_n^r, \omega_n^r, \phi_n^r\}_{r=c,s})\}_{n=1}^N$ which become a complete basis when $N \rightarrow \infty$, but is “chopped” at finite N and usually subjected to some degree of (momentarily affixed) randomization in frequencies and phases. Optimization is then typically carried out on the expansion coefficients and the control landscape reduces further to $F = F(c^c, c^s)$. Similarly to the toy problem above, a caveat hereto is that such a randomized parametrization may also inadvertently induce local traps not inherent to the full parametrized control problem. A way of escaping such “artificial” traps (Dressed Chopped Random Basis (DCRAB)) was proposed in Ref. [242] and is fairly straightforward: when the optimizer converges to an optimum at iteration k , we simply let $u_{\text{ref}} \leftarrow u(t; \theta^{(k)})$, zero the parametric basis expansion by setting $c^c = c^s = \mathbf{0}$, re-draw the randomization parameters, and resume the optimization in the new-drawn basis. Each of these resets constitute a kind of super-iteration that can be continued until the optimum is sufficiently basis-invariant and thus approximately a true optimum in the full parametrization landscape $F(\theta)$ — which may itself be an artificial minimum with respect to $F[u(t)]$ but that is an ever-present possibility by construction.

We have now explored how the control itself is amenable to parametrizations similar to those employed in seeding. The hope is then that both of these are able to effectively capture the optimal control strategies. One important consideration remains, namely that of handling control constraints.

3.6 COST FUNCTIONALS AND HANDLING CONSTRAINTS

It is nearly always necessary to handle optimizational constraints for a variety of theoretical and experimental reasons. There are several different ways to achieve this depending on the particular context. This section briefly reviews some of these mechanisms.

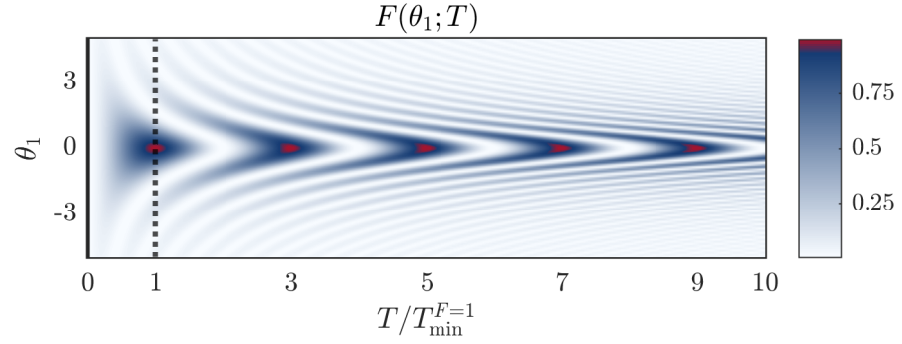


Figure 3.11: Control parametrization example ($M = 1$) where $u = u(t; \theta_1) = \theta_1$ allows us to map out the control landscape $F(\theta_1; T)$ for different T in units of $T_{\min}^{F=1} = \pi$ and $T = T_{\min}^{F=1}$ is shown with the vertical dotted line. Vertical slices correspond to the one-dimensional renditions of the control landscape. The color scale is sensitive to near-unit fidelities.

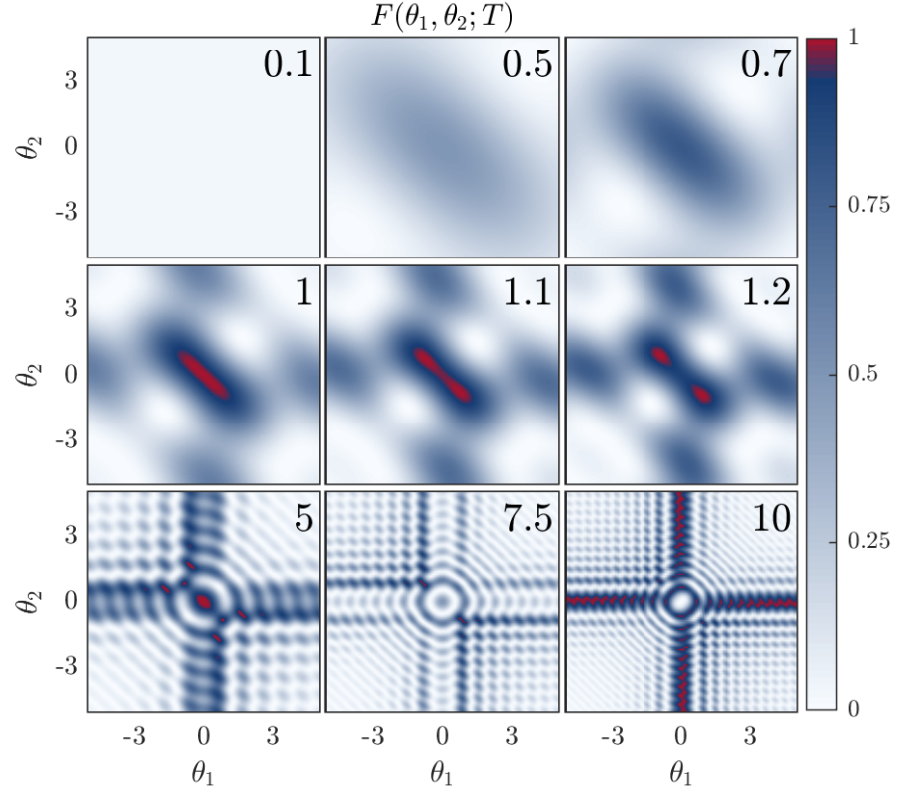


Figure 3.12: Same as Fig. 3.11 but with $M = 2$, $u = u(t; \theta_1, \theta_2) = \theta_1 \cdot I(t < T/2) + \theta_2 \cdot I(t \geq T/2)$. The control landscape $F(\theta_1, \theta_2; T)$ is shown for different T (shown in number labels) in units of $T_{\min}^{F=1} = \pi$. Note that $F(\theta_1, \theta_2 = \theta_1; T) = F(\theta_1; T)$ of Fig. 3.11.

Let us begin by introducing a superficial change in convention [214] by turning the maximization problem of F into the equivalent minimization problem of $-F$,

$$J_F[u(t)] = \frac{1}{2}(1 - F), \quad (3.21)$$

which additionally contains both a cosmetic scaling and offset. The former is in anticipation of a factor 2 when calculating derivatives and the latter is so that $0 \leq J_F$. In minimization nomenclature J_F is called a *cost* functional. The local maxima of F are the local minima of J_F and vice versa — the control landscape is “inverted” and we are now seeking troughs instead of peaks¹⁰. This is purely conventional and our analyses up to now are unchanged. We will use whatever convention makes sense for the context from hereon out.

Recall in our discussions that the task-encoding functional is a malleable entity. We can therefore compose it of any number of independent cost functionals,

$$J = \sum_i J_i. \quad (3.22)$$

Each cost term J_i — including the fidelity cost J_F and in fact any other imaginable control task cf. Sec. 3.1 — represents either a constraint or preference towards or against particular regions in the control landscape. A few examples are as follows.

Quantum equation of motion — Introduces a “penalty” for not following the equation of motion [1], $(i\partial_t - \hat{H}_{\text{gp}}(u)) |\psi\rangle = 0$,

$$J_{\text{gp}} = \Re \int_0^T \langle \chi(t) | (i\partial_t - \hat{H}_{\text{gp}}(u)) | \psi(t) \rangle dt. \quad (3.23)$$

We consider the GP case, $\hat{H}_{\text{gp}} = \hat{T} + \hat{V} + g|\psi|^2$, since it also covers the usual Schrödinger equation by taking $g = 0$. This type of cost is for a special use only as we shall see in Ch. 4. It has no weighting factor because it is always analytically zeroed exactly and it introduces an auxiliary state $|\chi(t)\rangle$ following an equation of motion that is not necessarily the same as for $|\psi(t)\rangle$. Similar costs can be defined for other equations of motion e.g. $(i\dot{\rho} - [\hat{H}, \rho]) = 0$ which leads to an auxiliary density matrix $\sigma(t)$.

Soft bounds — Introduces a penalty to control values outside given boundaries [1], $u_{\min} \leq u(t) \leq u_{\max}$,

$$J_\sigma = \frac{\sigma}{2} \int_0^T \Theta(u_{\min} - u)(u - u_{\min})^2 + \Theta(u - u_{\max})(u - u_{\max})^2 dt, \quad (3.24)$$

¹⁰ Instead of “climbing” peaks we are “rolling” downhill with an “equation of motion” dictated by local derivatives.

where Θ is the Heaviside step function and the weighting factor σ is typically of the order $10^3 - 10^4$ to heavily penalize controls outside the bounds. The bounds are nevertheless “soft” in the sense that exceeding them gives a smooth increase in cost. It is preferable to “hard bounds” (i.e. using the step functions only) due to differentiability. The price paid is that very small excursions over the boundaries are still permitted because of the squaring. This is of no real practical concern because one can then just contract the boundaries a bit as compensation.

Derivative regularization — Introduces a penalty to fast local oscillatory behavior [1],

$$J_\gamma = \frac{\gamma}{2} \int_0^T \dot{u}^2 dt \quad (3.25)$$

where γ is a weighting factor typically of order $10^{-8} - 10^{-4}$ and can optionally be scaled by the overall duration $\gamma \rightarrow \gamma/T$. This cost is used to shift preference towards smoother controls that are more consistent with experimental electronic bandwidths. It is remarked that this cost does not *guarantee* the absence of very large local oscillations as long as they are few and short (such as the ramping of broad bangs). Nevertheless, based on experience, even a small γ can tame rather large, unwanted rapid fluctuations decently throughout the entire control. Often such local fluctuations are numerical artifacts that averages out to an effective value for the resulting motion.

Amplitude regularization — Introduces a penalty to larger control values [4],

$$J_\alpha = \frac{\alpha}{2} \int_0^T u^2 dt \quad (3.26)$$

where α is a weighting factor typically of order $10^{-8} - 10^{-4}$ and can optionally be scaled by the overall duration $\alpha \rightarrow \alpha/T$. If u corresponds to e.g. a laser intensity, this cost is used to shift preference towards lower overall control values (and thus laser fluence) and minimize operational energy consumption.

Population discrimination — Introduces a penalty to populating a certain state $|\phi\rangle$,

$$J_\kappa = \frac{\kappa}{2} \int_0^T |\langle \phi | \psi(t) \rangle|^2 dt, \quad (3.27)$$

where κ is a weighting factor. Here, e.g., $|\phi\rangle = |E_n^{(0)}\rangle$ could be a bare-Hamiltonian eigenstate that is leaky or decoherence prone [243] to states outside the simulated manifold. Instead of including these transitions in the model explicitly, we can instead add a term that circumvents the offending state in question.

This non-exhaustive list are all so-called running costs, i.e. the penalty is recorded and accumulated across the entire process duration [215]. This is opposed to terminal costs in which the penalty is recorded only at the end of the process, e.g. as for the fidelity cost. The price to pay for including any of these additional terms is that we must compute their variations and include their derivatives in calculating the search direction $p^{(k)}(t)$ or $\mathbf{p}^{(k)}$, depending on whether a control parametrization is used or not. As before, we defer these calculations to Ch. 4. Additionally, adjusting the weight hyperparameters appropriately can be non-trivial.

Introducing new cost terms is not the only way to handle constraints. Below are a few examples that follows different approaches.

Non-linear control transformation — Control boundary constraints can be achieved by a passing u through a differentiable monotonic nonlinear transformation [1],

$$g = g(u(t)), \text{ where } g(-\infty) = u_{\min} \text{ and } g(+\infty) = u_{\max}. \quad (3.28)$$

such that $\hat{H}(g(u))$ and u itself can then be unbounded. The non-linear transformation g could for example be a shifted and scaled sigmoid or arctan function. This requires a modification to the variational calculations and associated resulting derivatives through the chain-rule.

Direct projection and cutoff — Control boundary constraints in both real and frequency space can be enforced simply by for example projecting back into the feasible region or directly cutting off (zeroing) exceders, respectively,

$$u(t) \leftarrow u(t)I(u(t) \in b) + u_{\min}I(u(t) < b) + u_{\max}I(u(t) > b), \quad (3.29)$$

$$\tilde{u}(\omega) \leftarrow \tilde{u}(\omega)(1 - I(|\omega| > b_{\omega})). \quad (3.30)$$

Here, $b = [u_{\min}, u_{\max}]$ and $b_{\omega} = [0, \omega_{\max}]$ are bounds in real and frequency space and $\tilde{u}(\omega)$ the Fourier components of $u(t)$. In this context, $\in$ is a boolean membership test and the relational operators are boolean tests if a value is less (greater) than the smallest (largest) element in the interval. After each normal control update iteration these operations are applied as a post-processing step. This type of approach is obviously quite crude because it can essentially erase, at each application, some or all of the optimizer's hard work instead of working with it in harmony. It may also not be obvious how to apply these operations on parameterized controls. Nevertheless, it has been found to work well enough in certain cases [244].

Filter function — Electronic equipment has a finite bandwidth that can distort an input control signal $u(t)$ into a slightly modified and

smoother version $v(u(t))$ and the actually realized dynamics is due to $\hat{H}(v(u))$. The distortion is most significant and thus detrimental when the local switching time is fast. If for example u contains bangs, $v(u)$ will have similar bangs but with a finite rising time between the extremal values. The two are related by a filter function h (typically experimentally estimated) through the convolution $\star$ [165],

$$v(u(t)) = (h \star u)(t) = \int_0^t h(\tau) u(t - \tau) d\tau. \quad (3.31)$$

This requires a modification to the variational calculations and associated resulting derivatives through the chain-rule.

Control parametrization — In our discussion about control parametrization we have seen that certain basis expansions are amenable to handling constraints. For example, bandwidth limitations can be directly incorporated when expanding on Fourier basis simply by limiting the number of terms (each subsequent term is of an increasing harmonic frequency). This essentially corresponds to a cutoff frequency. This can work very well in experiment when compared to a piecewise constant basis expansion [62].

Algorithmic constraint handling — Certain optimization algorithms are built with sophisticated constraint handling in mind. In particular, so-called Interior-Point methods [4, 214, 245] work by incorporating the constraints at the level of the search direction. This type of approach is probably among the best since it has certain provable theoretical properties. Another example is a variant of the limited-memory BFGS search direction, L-BFGS-B [246] (the last “B” is for “bounds”), which can handle simple box constraints.

In conclusion, there are many different ways to handle control constraints. Some are very simple and can work decently well on average whereas others are more involved but also typically carries a degree of rigor and more attractive guarantees. Any number of these may be mixed and matched depending on the context and there is no clear one-size-fits-all answer.

3.7 FURTHER DEVELOPMENTS AND CONCEPTS

We have now established most of the core ideas for the optimization. There are, however, still a variety of more advanced topics that are very useful or necessary to consider. A few of them are as follows.

3.7.1 Robustness

Specifying control constraints is vital when the purpose of devising optimal controls is to deploy them in particular experimental setups.

For a variety of reasons, however, it may happen that certain parts of the setup or experimental protocol are not completely specifiable and thus yet hampering the achievable performance. Perhaps the system parameters are characterized by a distribution (e.g. fabrication-induced inhomogeneous artificial atom qubits), are drifting or fluctuating in time (e.g. control laser intensity), are different run-to-run (e.g. number of condensed BEC atoms), or some other stochastic or systematic process that may interfere with the exact predictability of the system. As far as high-fidelity operations are concerned, this can severely limit the potency of model-based optimal controls if these effects are not accounted for. After all, any optimal control conceived this way is only as good as the underlying physical modeling. We discuss this further in the Sec. 3.8.4.

One way to endow optimal controls with noise resilience is through so-called *ensemble* optimization [62, 244]. The system Hamiltonian can be written as $\hat{H} = \hat{H}(u; \mathbf{x})$ where $u(t)$ is the control (laser intensity, laser position, ...) and $\mathbf{x}$ is the vector of uncontrollable system parameters (coupling constants, number of atoms, ...). We can then write a cost for an ensemble of similar systems with slight variations,

$$\bar{J}_F = \sum_{l=1}^L p(x_l) \cdot J_F[u(t); \mathbf{x}_l], \quad (3.32)$$

where each system parameter in the vector $\mathbf{x}_l$ is sampled either at random or regular intervals from some characteristic probability distribution $p(\mathbf{x})$, e.g. Gaussian or bounded uniform. Optimizing $\bar{J}_F$, the weighted mean of different realizations, will then produce $\bar{u}^*(t)$ that produces good results with respect to the overall ensemble. That is, it is inherently robust towards the *a priori* specified parameter fluctuations, equivalently located in simultaneous troughs in the L different control landscape instantiations $J_F[u(t); \mathbf{x}_l] \approx 0$.

Similarly we could add characteristic (possibly time-dependent and sampled from run-to-run) noise terms $\delta u(t)$ and/or $\delta \mathbf{x}(t)$ to also stabilize against these effects. Optimal controls hereof will then predominantly be in troughs of control space that are relatively broad such that small perturbations are less detrimental $J_F[u(t) + \delta u(t); \mathbf{x}_l + \delta \mathbf{x}(t)] \approx 0$.

Each term in Eq. (3.32) requires evaluation of its corresponding derivatives and this methodology can then obviously quite quickly become an expensive combinatorial explosion. Outside of additional numerical effort it is also clear that the ensemble control problem itself is much more difficult than the original one. The methodology should therefore be used sparingly and purposefully. For example, if the effects are sufficiently small, one could initially optimize the most likely single realization and, upon converging, use that $u^*(t)$ as a seed in including the full ensemble to also make it robust.

3.7.2 Homotopies

In mathematical topology, two functions are said to be *homotopic* if one can be “continuously deformed” into the other and the deformation itself is called a *homotopy*. We may appropriate this notion in the context of optimization/optimal control as follows¹¹ [214, 216, 245].

Store as elements in the generalized vector $\mathbf{h}$ a subset of all the parameters that defines an optimization task in the discrete numerical setting. This could for example include the initial and target states, temporal discretization δt and total duration $T = (N_t - 1)\delta t$, spatial discretization δx and number of points n_x , Hamiltonian parameters, number of particles or sites, number of optimization parameters or basis functions, certain numerical truncation parameters, and so on. That is, $\mathbf{h} = (|\psi_{\text{tgt}}\rangle, |\psi_{\text{ini}}\rangle, \delta t, T, \dots)$ and we call each element hereof a homotopy parameter. We may then define a homotopy (or continuation) between two optimization tasks A and B , individually defined by a certain choice $\mathbf{h}_{A(B)}$, as a sequence of “sufficiently quasi-continuous” homotopy parameters, $\{\mathbf{h}^{(k)}\}_{k=1}^K$, such that

$$\mathbf{h}_A = \mathbf{h}^{(1)} \rightarrow \mathbf{h}^{(2)} \rightarrow \dots \rightarrow \mathbf{h}^{(K)} = \mathbf{h}_B. \quad (3.33)$$

The general idea is then that a solution to the problem defined by $\mathbf{h}^{(k)}$ is also at least an approximate solution to $\mathbf{h}^{(k+1)} \approx \mathbf{h}^{(k)} + \delta\mathbf{h}$ where $\delta\mathbf{h}$ is a “sufficiently small” change in parameters. The upshot is that we may significantly accelerate — under appropriate and careful considerations — the acquisition of solutions to a “hard” problem $\mathbf{h}_B$ by solving a sequence of “easy” problems starting from $\mathbf{h}_A$, handing over the interim solutions. Here, easy and hard (quotation marks are dropped from hereon) can be understood in both terms of computational effort and/or inherent landscape complexity as discussed later in Ch. 8. We could also augment this formulation to append at the end $\mathbf{h}_B \approx \mathbf{h}_C$ where $\mathbf{h}_C$ is representative of the “true” optimization task in the limit of all homotopy parameters being sent to their natural fine-grained limits, e.g., $\delta t \rightarrow 0$. This is in fact always an implicit assumption in *any* optimization if one is to expect meaningful results.

As an illustrative and generally applicable example consider a single homotopy parameter in the temporal discretization $\mathbf{h} = (\delta t)$. To obtain physically meaningful results and similar, one must choose a $0 < \delta t \leq \delta t_{\text{max}}$ small enough to adequately resolve the “true” dynamics in the sense of $\mathbf{h}_C = (\delta t \rightarrow 0)_C$. For some particular choice consistent with this requirement, say $\tilde{\delta t}$, this defines $\mathbf{h}_B = (\tilde{\delta t})_B$. This original problem is hard, because it potentially requires many N_t steps to integrate the equations of motion up to a fixed T . Nearly always, however, solutions that are not quite entirely physically meaningful ($\delta t \geq \delta t_{\text{max}}$) may yet capture essential features of the process while being much more easy

¹¹ Incidentally we have already seen a similar formulation in discussing the optimal control strategies in Sec. 3.2.1.

to reach because of fewer time integration steps. That is, it may be appropriate to define a homotopy sequence, e.g.

$$\mathbf{h}_A = (4\tilde{\delta}t) \rightarrow (2\tilde{\delta}t) \rightarrow (\tilde{\delta}t) = \mathbf{h}_B, \quad (3.34)$$

corresponding to an interpolation between coarse and fine grained resolution. On that note, the control itself must be interpolated upon changing to a new time grid and this is best done piecewise when the δt -homotopy is scaled by an integer p corresponding to $\hat{\mathcal{U}}(\delta t) \approx \hat{\mathcal{U}}(\delta t/p)^p$. The caveat is that we should not spend too long solving the coarse grained problem since the details thereof are not relevant for the original problem.

The primary concerns for effectively utilizing homotopy techniques thus lie in identifying (1) a relevant set of homotopy parameters $\mathbf{h}$, (2) defining an appropriate sequence of explicit values for these, and (3) striking a balance between time spent on the easier problems $\{\mathbf{h}^{(k)}\}_k^{K-1}$ and the original hard problem $\mathbf{h}_B = \mathbf{h}^{(K)}$. Although this heuristic does not strictly guarantee an advantage it works very well in practice. In Refs. [2],[4], we took homotopy sequences in $\{n_x, \delta t\} = \{32, 5 \cdot 10^{-4}\} \rightarrow \{64, 1 \cdot 10^{-4}\} \rightarrow \{64, 1.2 \cdot 10^{-5}\}$ and $\{\delta t\} = \{0.1\} \rightarrow \{0.05\} \rightarrow \{0.025\}$, respectively, which roughly reduced the time scale for optimization from weeks to days.

3.7.3 Imaginary bounds

The systems under here consideration are trapping geometries without periodic boundaries. That is, sites at 1 and N_{sites} and spatial wave function amplitudes at index 1 and n_x are not coupled to each other. For the latter, however, time-evolution is performed with the split-step (Trotter) Fourier integration technique which introduces such unwanted coupling. The optimizer is only privy to the implemented dynamics and cost functional and may therefore discover solutions that are mediated by this coupling — colloquially, the optimized wave function can travel through the left boundary and reappear at the right boundary. This clearly unphysical behavior must therefore be disallowed somehow.

One way to do this is to add an imaginary potential, $V_{\text{im}}(x) = -icV_{\text{bounds}}(x)$, of strength c and boundary function, e.g.,

$$V_{\text{bounds}}(x) = \Theta(\tilde{x}_{\min} - x)(x - \tilde{x}_{\min})^2 + \Theta(x - \tilde{x}_{\max})(x - \tilde{x}_{\max})^2. \quad (3.35)$$

This resembles the control soft-bounds in Eq. (3.24) and the squaring is for the same reason as there. Here $\tilde{x}_{\min} = x_{\min} + \Delta x$ and $\tilde{x}_{\max} = x_{\max} - \Delta x$ are “effective” limits for the problem, contracted slightly from the full numerical limits. Beyond these the wave function is attenuated without re-normalization in each time step since $\exp(-iV_{\text{im}}\delta t) = \exp(-cV_{\text{bounds}}\delta t)$ which is unit inside the effective

limits and exponentially decaying outside [1]. The attenuated wave function will result in low or vanishing fidelity because the norm has reduced and hence the optimizer does not explore these unphysical solutions.

3.8 PERSPECTIVES TO OTHER AREAS OF QUANTUM CONTROL

The base premise of this thesis is to control quantum systems and therefore decidedly belongs to the broad field of *quantum control*. We also wish to do so in the best possible, or optimal, way and hence the moniker quantum *optimal* control. As alluded to in the introduction, however, this is yet an area with a large variety of different approaches and goals as for example laid out in Ref. [52] which is due to several prominent figures within the community. Quantum optimal control can therefore mean different things to different people which is what prompted the decorated specification of my small corner in the introduction. In this section we briefly touch and reflect upon a few other alternatives to the herein presented research questions and methodologies.

3.8.1 Controllability

Controllability is a part of quantum optimal control that seeks to answer if any arbitrary state (or set of states) can be dynamically connected to any other arbitrary state (or set of states), equivalently any unitary operator can be obtained from the identity. If the answer is in the affirmative the system is said to be *controllable* [47, 216, 247]. Note that this is a much broader question than single state transfer problems where we are only concerned with connecting two specific states. It is also a different *type* of question since it does not principally concern actually designing the optimal controls. Controllability of closed finite-dimensional bilinear systems is relatively well-understood whereas the corresponding infinite-dimensional and open-system studies are much more challenging and are active areas of research [52].

The typical controllability test for a closed N -level bilinear system is based on the so-called Lie algebra rank condition [47]. In short, the Lie algebra,

$$\text{Lie}\{i\hat{H}_0, i\hat{H}_1, i\hat{H}_2, \dots\}, \quad (3.36)$$

is the subspace formed by all repeated commutators between the drift ($\hat{H}_0$) and control Hamiltonians ($\hat{H}_{i=1,2,\dots}$) and the rank is given by the number of resulting elements. It can then be shown that the system is controllable if and only if the Lie rank equals N^2 . Qualitatively we can understand this as follows. The less commutative two operators are, the less similar they are and the more different their induced dy-

namics will be¹². That is, each Hamiltonian is responsible for driving the system along one direction and the more distinct or “orthogonal” these available directions are, the more easy it is to control the state arbitrarily. Incidentally, similar ideas apply in establishing the potency of a quantum simulator in e.g. Ref. [16]. There, Lloyd gives a vivid classical analogy of trying to drive a car from A to B and the route taken is given by a sequence of applied Hamiltonians at ones disposal. We could extend this analogy to cover an example of non-controllability — if the destination B is not in the plane of the car but some distance above, it will not be able to reach it and the system is non-controllable. Yet, for a futuristic car with flying capabilities, correspondingly a new orthogonal Hamiltonian, the situation may again be controllable.

Nevertheless, even if the system is found to be controllable, this does procure the actual optimal controls or the minimal duration T_{ctrl} for controllability. Additionally, we will for virtually all pairs of initial and target states have $T_{\text{min}}^{|\psi_{\text{ini}}\rangle \rightarrow |\psi_{\text{tgt}}\rangle} \ll T_{\text{ctrl}}$ and in many practical applications it is not important that the full system is controllable so long as the particular transfer(s) we desire can be performed.

3.8.2 Minimal durations and quantum speed limits

The minimal duration T_{min} necessary to perform a transfer between two distinguishable [249] quantum states is fundamentally bounded by the laws of quantum mechanics. A surprisingly simple interpretation can be found in Heisenberg’s uncertainty relations for time and energy $\Delta\hat{H}\Delta t \geq \hbar/2$. Compared to the uncertainty relations for position and momentum $\Delta\hat{x}\Delta\hat{p}_x \geq \hbar/2$ — which is a statement on the simultaneous measurement precision between conjugate variables based on their dispersive canonical commutation relations [250] — the time-energy variant cannot be attributed the same understanding most simply because time is not an observable in quantum mechanics. It instead expresses an intrinsic time scale Δt for unitary quantum dynamics [250] by $\Delta t \sim \hbar/\Delta\hat{H}$. That is, there is a fundamental limit to how fast a state can change or react when the energy is finite. The energy dispersion $\Delta\hat{H}$ should then be maximized to minimize Δt . We remark in passing that adiabatic processes work exactly because of this: to ensure that the state remains very close to a particular instantaneous eigenstate (small ΔH) one must vary the Hamiltonian sufficiently slowly (large Δt).

This interpretation was initially due to Mandelstam and Tamm [251] who also proved a bound $T_{\text{qsl}}^{\text{MT}} = \pi\hbar/2\Delta\hat{H}$ for the *quantum speed limit* time which is defined as the minimal time necessary for an initial state to evolve into an orthogonal state [250], $\langle\psi(T_{\text{qsl}})|\psi(0)\rangle = 0$ — which

¹² This is also one of the main conceptual ideas in bang-bang structures. Incidentally, in a slightly different packaging, bang-bang has been proposed to be a universal feature for bilinear optimal controls [248].

for obvious reasons is therefore also called the orthogonalization time [252]. More recently Margolus and Levitin [252] provided a different bound $T_{\text{qsl}}^{\text{ML}} = \pi\hbar/2 \langle H \rangle$ and these were later combined and proved to be tight [253]

$$T_{\text{qsl}} = \max\{T_{\text{qsl}}^{\text{MT}}, T_{\text{qsl}}^{\text{ML}}\} = \frac{\pi\hbar}{2} \max\{(\Delta\hat{H})^{-1}, (\langle\hat{H}\rangle)^{-1}\}. \quad (3.37)$$

Unfortunately this unified bound assumes a constant Hamiltonian and that is at odds with the desire to manipulate the dynamics with a control $u(t)$. Extensions to the time-dependent case and similar analytical studies (e.g. geometric methods and non-unitary dynamics) is now a whole research topic in itself (see [250] and references therein).

The work presented in this thesis takes a different approach, namely by establishing bounds based on empirical discoveries (cf. Fig. 3.4) through the framework developed in this chapter. It is remarked that our discussions elsewhere have concerned T_{min} rather than T_{qsl} since the state transfer formulation does not *necessarily* have to be between orthogonal states and $T_{\text{min}} \leq T_{\text{qsl}}$ always [222]. Quite often the two durations are conflated and used synonymously in the literature and can indeed coincide, but they are, strictly speaking, not the same thing. Further to the confusion is that T_{qsl} is often referred to as *the* quantum speed limit although it is a time and not a speed.

Regardless of these caveats, T_{min} is bounded by the same fundamental mechanism as T_{qsl} — the Heisenberg uncertainty relation — and the dynamical time scale interpretation remains valid. Specifically, the fastest instantaneous motion in Hilbert space is achieved when populating several of the instantaneous energy eigenstates to maximize the energy spread. In most instances this is, interestingly, at odds with the initial and target states since typically at least one, and often both, are individual energy eigenstates with vanishing energy variance by definition. Assuming both are eigenstates, this implies that the system should undergo an initial rapid excitation into broad parts of the spectrum which must then tighten back up nearing the end of the transfer. This behavior is indeed routinely observed in practice when operating near T_{min}^F .

3.8.3 Coherent control

A flavor of quantum control is so-called *coherent control* [51, 52] which was conceived for chemical problems in the 1980's and as such seeks to solve the same type of control problems. Coherent control can therefore, to my limited understanding, either be considered its own apparatus or a part of quantum optimal control. While such semanticism may or may not be warranted, the following working distinction [254] may be useful at least with respect to the flavor of optimal control presented in this thesis.

Coherent control is a “bottom-up” approach where controls are selectively constructed from analytical considerations. In particular, careful interference between multiple quantum pathways may suppress or enhance the final state. Optimal control is, on the other hand, a “top-down” approach where we apply quite general algorithms to extract optimal solutions that extremize an encoding functional. Importantly, we do not specify the physical mechanisms responsible for any given transfer process (although we can of course still use appropriate seeding and parametrization if we have domain knowledge). Upon obtaining an optimal control we may then attempt to uncover why exactly it works and these may indeed coincide with solutions produced from coherent control.

As an instructive example of these differences, consider a 3-level Λ -system spanned by states $\{|n_1\rangle, |n_2\rangle, |n_3\rangle\}$. The goal is to perform the transfer $|n_1\rangle \rightarrow |n_3\rangle$. The states $|n_1\rangle$ and $|n_3\rangle$ are coupled to $|n_2\rangle$ by individual addressing lasers, but not to each other, whereby $|n_2\rangle$ acts as an intermediate state. The added complication is that $|n_2\rangle$ must remain unpopulated during the transfer — otherwise a trivial solution is just to drive the $|n_1\rangle \rightarrow |n_2\rangle$ with one laser and then $|n_2\rangle \rightarrow |n_3\rangle$ with the second laser. In the coherent control arena we can solve this problem by analytically diagonalizing the Hamiltonian, applying eigenstate and normalization conditions, and from there infer a Stimulated Raman adiabatic passage (STIRAP) protocol [51, 255]. Using instead quantum optimal control, we would optimize a cost functional $J = J_F + J_\kappa$ where J_F encodes the state transfer $|n_1\rangle \rightarrow |n_3\rangle$ and J_κ that populating the $|n_2\rangle$ state is not allowed (as defined in Sec. 3.6). If desired, the laser pulse shapes could be parametrized directly as Gaussians [243, 256]. In this way, we have two completely different approaches that solves the problem on their own merits — one where the causal mechanism is imposed and one where it is empirically and independently discovered.

3.8.4 The Role of Open-Loop Derivative-Based Optimization

Model-based open-loop methodologies produce controls that are optimal with respect to some chosen encoding functional $J[u(t)]$. If the study is purely theoretical and ideal conditions are assumed then the same controls will yield the same performance in the corresponding ideal experimental setup. Ideal experiments are nevertheless elusive. This prompts significant efforts into the theoretical developments of methodological extensions such as in the previous sections which allow us to account for a very wide variety of constraints and criteria. Very often a number of these are included even in still purely theoretical studies with only a hypothetical experiment in mind.

When the theoretically acquired controls are intended for deployment into a particular real-world experimental apparatus, these con-

siderations necessarily become much more vital. Clearly, an optimal control obtained this way can only be expected to be as good as the underlying model of the experimental reality¹³. This fact will nearly always lead to a performance degradation, $F^{\text{experiment}}[u^*] \lesssim F^{\text{theory}}[u^*]$, which may or may not be acceptable in the context of a particular application. This thesis is primarily geared towards the high-fidelity limit operating as close as possible to the shortest duration fundamentally allowed by quantum mechanics with a *modus operandi* rooted in derivative-based local optimization. Unfortunately this limit is also likely to be the most inherently sensitive to the performance delta and the methodology itself — although quite possibly the most high-performant theoretical tool as our disposal — potentially much more complicated than the alternatives. This becomes especially true with increasing system complexity, e.g. in the quantum many-body regime [4, 73, 74]. For example, direct search or global optimizers are much more amenable to the closed-loop setting which is attractive because the “encoding functional” then *is* the experimental reality by definition. Further, in this setting, the by far most difficult technical aspects of an open-loop approach — implementing equations of motion and potentially control derivatives — is completely circumvented and the optimization algorithmic logic can readily be outsourced to one of the many black-box software packages that work well off-the-shelf. This significantly reduces the barrier of entry in terms of programming proficiency and provides a comfortable scaffolding for e.g. experimental groups who are not necessarily domain experts in optimization or high-performance computing.

The question is, then, if open-loop derivative-based local optimization is a longterm viable strategy. In particular, if any substantial edge gained is ultimately washed out in experiment then the significant associated additional efforts may indeed appear redundant at best and a waste of time and resources at worst. Although it is well beyond this thesis to give definitive answers to this question I do not believe that the future is quite as bleak as the preceding paragraph might suggest. There are several promising and positive outlooks in this regard as follows.

A boon to open-loop derivative-based control is its high efficiency in uncovering optimal control strategies [4, 5], i.e. the general solution structures and associated physical processes that characterize a problem. In this sense, the physical modeling is only required to be sufficiently good enough that the optimal *strategy* is overall physically representative. Thus, it is of much less importance if any individual optimal control is degraded in experiment or not which is otherwise one of the main arguments against the approach. This type of insight is very valuable because it can lead to winning seeding strategies and control parametrizations which entail significantly reduced and

¹³ *Garbage in, garbage out*, in computer science terminology.

restricted search spaces. Subsequent to these identifications, one could then draw upon the strengths of closed-loop methodologies as described above and densely probe the control region selectively rather than the full space on realistic timescales and thus empirically account for, or even calibrate, potential model uncertainties. This could lead to a meta-iterative approach in which a comprehensive ecosystem of methodologies work in unison [163, 257, 258], rather than participating in a counter-productive inimical competition, drawing on their individual capabilities to minimize the performance gap between theory and experiment. There is no free lunch [259] — but if every methodology chips in, the lunch may indeed be more affordable. Yet, it is also worth noting that open-loop control by itself can, in fact, achieve significant experimental fidelities even in large Hilbert space processes for which no intuitive solution is known [78] if the experimental reality is well-modeled.

These compounded observations seem to suggest that open-loop derivative-based control may indeed maintain relevance and we shall without any further reservation explore the details of this approach.

*Ak ja,
Livet er svært -
men Matematik er sværere.*

— Storm. P.

THE elements of quantum optimal control are laid out in the previous chapter, but one central ingredient is missing to complete the description. Namely, a set of working equations that allows us to actually optimize the encoding functionals with derivative-based local optimization which is our approach of choice. This chapter derives these equations, tours some technical details of this setting, and discusses the actual update steps in Eqs. (3.9) and (3.17) which will both end up on the numerically amenable vectorial form $\alpha^{(k)} \mathbf{p}^{(k)}$ after discretization.

The chapter is structured as follows.

Section 4.1 discusses the importance of exact derivatives and outlines the considerations in choosing a search direction and step size.

Section 4.2 lays out the general structure for temporal discretization and introduces two possible lines of derivation for analytical derivatives (optimize-then-discretize and discretize-then-optimize) in anticipation of delving into their respective *tour de force* calculations.

Section 4.3 follows the optimize-then-discretize path and calculates derivatives using the calculus of variations.

Section 4.4 follows the discretize-then-optimize path and calculates derivatives using ordinary partial derivatives and explicitly takes into account specific numerical implementations.

Section 4.5 summarizes and compares aspects of the two paths and argues preference to discretize-then-optimize.

Section 4.6 reflects and comments on a suite of named algorithmic approaches and attempts to delineate a variety of emergent acronyms. Discussion of these *quantum optimal control algorithms* (GRAPE, GROUP, GOAT, CRAB, and KROTOV) is purposefully saved for last.

4.1 LOCAL SEARCHING — TAKING THE STEP

Accurate and efficiently calculable derivatives are vital requirements for the success of local search methodologies. Inexact derivatives can

lead us irrecoverably astray and limit the acquisition of high-fidelity results whereas crippling evaluation times will prevent reaching any kind of optimized solution in finite time. Appealing yet again to a simple landscape analogy, it would be similarly quite challenging to expeditiously climb or descend a mountain armed only with a faulty compass that is also slow to read off. These criteria are valid across all optimization tasks, but tighten in the high-fidelity regime and with increasing system complexity and Hilbert space size where every single optimization cycle counts.

For the presentation given in this section we will assume that temporal discretization has taken place as extensively detailed in the later following sections. All we need for now is that the update rules read

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha^{(k)} \mathbf{p}_{\mathbf{u}}^{(k)}, \quad \boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \alpha^{(k)} \mathbf{p}_{\boldsymbol{\theta}}^{(k)}, \quad (4.1)$$

depending on whether the control is $u(t)$ or parametrized $u(t; \boldsymbol{\theta})$ by $\boldsymbol{\theta}$ prior to temporal discretization. The former turns out to just be a special case of the latter in the discrete setting. The distinction is therefore not particularly important and we will use $\mathbf{u}$ or $\boldsymbol{\theta}$ somewhat interchangeably and not explicitly label $\mathbf{p}^{(k)}$.

What *is* important, however, is the search direction $\mathbf{p}^{(k)}$ itself — the compass needle in which we trust to faithfully illuminate the fidelity peaks or cost troughs¹. After discretization we have a finite number of parameters which means we are now essentially applying nonlinear optimization to quantum physics. Nonlinear optimization [214, 245, 260–262], or programming, is a vast piece of mathematical machinery with a rich history and contains layers of sophistication, tricks, and offers a multitude of algorithmic variations. We inherit all these trappings “for free” and it is not the purpose of this thesis to give a detailed mathematical account of all these important elements. These can be found in standard textbooks on the subject and I have always found Ref. [214] useful. Instead we simply highlight and discuss the most central aspects here. In the later following sections we then focus on the “quantum-part”, namely the equations of motion which plays a central role in calculating control derivatives of the cost J — the gradient ∇J and possibly the Hessian $\nabla^2 J$ — which are necessary to form the search direction $\mathbf{p}^{(k)}$.

4.1.1 Search Directions

Let us explore a variety of possible compass needles. Most of these are inspired by Taylor’s theorem [214]. To simplify notation we will write e.g. $\nabla J^{(k)} = \nabla_{\mathbf{u}} J(\mathbf{u}^{(k)})$ keeping *in mente* that we are referring to control derivatives evaluated locally at $\mathbf{u}^{(k)}$. The derivatives themselves

¹ The sign convention chosen in maximizing or minimizing equivalent functions manifests in a sign change of search directions, $\mathbf{p}^{(k)} \rightarrow -\mathbf{p}^{(k)}$.

are assumed to be readily calculable by some unspecified description. For later reference, the lower bound on the computational cost of calculating the gradient is roughly 2 full-duration propagations i.e. $2N_t$ individual propagation where N_t is the number of discrete time indices. This lower bound is only achieved for a certain few methodologies (cf. Sec. 4.5). A similarly conditioned lower bound for the Hessian computational cost roughly requires an additional $N_t^2/2$ propagations.

Steepest Descent Direction

The simplest choice of direction is the direction of steepest descent $\mathbf{p}_{\text{steepest}}^{(k)} = -\nabla J^{(k)}$. The gradient is always orthogonal to the local contour line and points in the direction of steepest increase in the objective function and the negation thus in the direction of steepest decrease. Any direction with a negative *directional derivative*², $\langle \nabla J^{(k)}, \mathbf{p}^{(k)} \rangle < 0$, where $\langle \cdot, \cdot \rangle$ is an inner product is called a descent direction. Although steepest descent sounds like a promising strategy it comes with several unattractive theoretical properties such as poor convergence rates (linear) [214] and a propensity to produce zig-zag patterns in optimization space when a “straight line down the zig-zag-center” would be preferable.

Newton Direction

A theoretically stronger search direction is the Newton direction $\mathbf{p}_{\text{Newton}}^{(k)} = -(\nabla^2 J^{(k)})^{-1} \nabla J^{(k)}$ which assumes a locally quadratic model and requires evaluation of both gradient and Hessian. The Newton direction thus incorporates curvature information. As it approaches the optimum where the quadratic model is a good approximation it enjoys a quadratic convergence rate [214] and $\alpha^{(k)} \rightarrow 1$. It is also more robust (scale invariant) towards poorly scaled problems. Problem scaling is discussed separately in a later section. There are two main issues with the Newton direction, however. First, far from the optimum, the Hessian may not be invertible and the direction thus not defined (although there are certain approximation tricks to alleviate this issue to some extent [263]). Second, calculating the Hessian in quantum contexts is typically very expensive and may not be worth the additional computational overhead.

Quasi-Newton Directions

For these reasons we instead in practice typically construct an approximation to the Newton direction that is free of both problems. This class of search directions are collectively referred to quasi-Newton

² Later we will use directional derivatives to actually define the notion of gradients for any normed vector space.

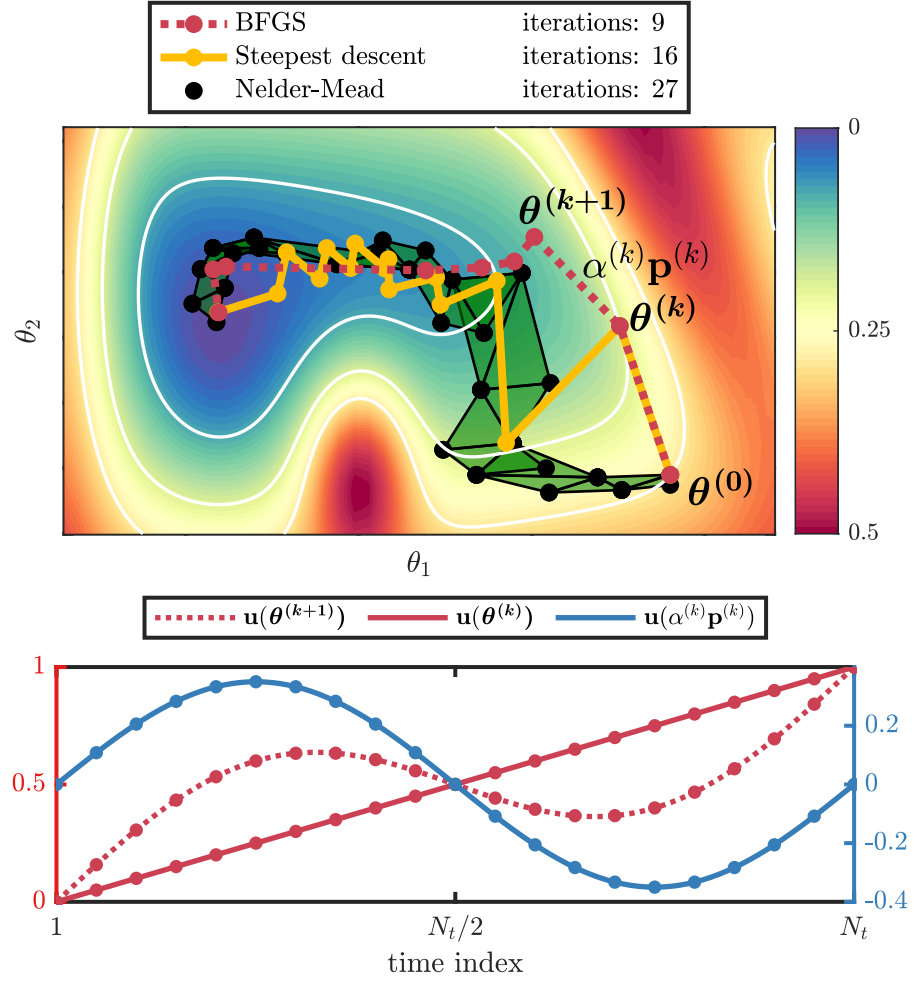


Figure 4.1: Top: Numerical example of following $\mathbf{p}_{\text{BFGS}}^{(k)}$ or $\mathbf{p}_{\text{steepest}}^{(k)}$ or using a Nelder-Mead direct search for some synthetic cost function J . The control is generically parametrized $u(\theta_1, \theta_2)$. Bottom: Example of how the controls at iteration k and $k+1$ and associated step could look like after assembling the parametrization (left axis). Here the shorthand $u(\alpha^{(k)} \mathbf{p}^{(k)}) \sim u(\theta^{(k+1)}) - u(\theta^{(k)})$ (right axis) is understood as the update in the time-domain resulting from the update $\theta^{(k+1)} = \theta^{(k)} + \alpha^{(k)} \mathbf{p}^{(k)}$ in the parameter space.

directions. The most popular one is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) direction, $\mathbf{p}_{\text{BFGS}}^{(k)} \approx \mathbf{p}_{\text{Newton}}^{(k)}$, which is built iteratively from local gradients encountered along the optimization trajectory. Briefly, it is based on requiring certain appropriate properties for successive gradients in solving the so-called secant equation. Quasi-Newton methods come in several variants and the relevant recorded information may be stored in matrices or, more frugally, in limited-memory formats [214, 264]. Typically they are associated with super-linear but sub-quadratic convergence rates [214] and are widely considered the best of the “simple” standard search directions. Another honorable mention is the conjugate-gradient direction (also built from recorded gradients) but this seems to nearly always be outperformed by quasi-Newton-types in practice. In passing it is remarked that most standard descent directions can be written as $\mathbf{p}^{(k)} = P(\nabla J^{(k)})$ where P is a *preconditioner* [214] applied to the gradient and thus highlighting its central role.

Practical Concerns

As far as the quantum context is concerned, assembling a quasi-Newton direction incurs negligible additional overhead per iteration compared to just calculating the gradient and should therefore always be used over the steepest descent direction.

An illustrative numerical example of following the local $\mathbf{p}_{\text{BFGS}}^{(k)}$ and $\mathbf{p}_{\text{steepest}}^{(k)}$ directions is shown in Fig. 4.1. It also shows a Nelder-Mead direct search consisting of an $M + 1$ dimensional polytope/simplex (here triangles) trickling downhill. The landscape itself is synthetic but could correspond to a quantum task with e.g. a parametrized control $\mathbf{u}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^{M=2}$ as in Fig. 3.12. The steepest descent trajectory is plagued by the inconvenient zig-zag behavior as the direction is always orthogonal to the contour lines. The BFGS trajectory circumvents this undesirable feature due to the iteratively constructed curvature information. Note that the first BFGS step is in the steepest descent direction as no information has yet been recorded. The number of iterations needed for BFGS is significantly lower than for steepest descent and Nelder-Mead. In practice, optimization is typically performed in much higher dimensional and more challenging control landscapes and these discrepancies are always strongly (nonlinearly) enhanced. One feature not captured by the example is the slow convergence of steepest descent near the optimum but this effect is, to usually even greater prominence, also strongly enhanced.

We can now appreciate the importance of exact derivatives more quantitatively. Any inexactness, due to willful approximations or otherwise, corresponds to geometrical misalignment with respect to the intended search direction. The steepest descent direction is subordinate to begin with, but an inexact gradient will no longer truly point

in the direction of steepest function decrease. The substantial effort in calculating the Newton direction will be voided by inexact gradients and Hessians making it now a completely unfeasible choice. The consequences for quasi-Newton-type directions such as [BFGS](#), otherwise usually the best choice, are affected in a more sinister way. Since these are built from successive gradients, the individual local errors are compounded which completely scrambles the Hessian approximation during optimization. Ameliorative countermeasures can be formed by periodically flushing and restarting the gradient history, but this introduces situational band-aid complexity and throws away what could and should have been valuable information. In conclusion, exact and efficiently calculable derivatives are very important.

Advanced Directions

The “simple” search directions mentioned in the above are, nevertheless, just scratching the surface of what is on offer within the nonlinear optimization literature. There are many additional developments, perhaps most notably the family of interior-point directions $\mathbf{p}_{\text{Interior-Point}}^{(k)}$ which builds off of $\mathbf{p}_{\text{BFGS}}^{(k)}$ or $\mathbf{p}_{\text{Newton}}^{(k)}$ and is capable of incorporating control constraints in a sophisticated and theoretically sound way³. The implementation details of these, however, quickly become extremely demanding [[214](#), [265](#)], much more so than most physicists, including myself, would realistically want or need to implement for themselves. Instead the typical scenario is to settle for the simpler choices above or to use some external implementation where one just has to supply the derivatives. In reflection of my own experiences I would have, given the timely opportunity, advocated for choosing the latter in developing the QENGINE C++ library [[1](#)]. That is, keeping the high-performance “quantum”-parts (dynamics, derivative calculations, overall wrapping and representation of the control task) but outsourcing the actual optimization logic to a separate high-performance implementation such as IPOPT [[265](#)] in much the same way we did for the project in Ref. [[4](#)]. These musings are not of crucial nature, however, since the optimizational performance of the QENGINE has been good in its applications and is also well-tested. Still, it is worth considering for the future in an effort to push the performance closer to its absolute limits.

4.1.2 *Line Searching*

Suppose we have picked out some appropriate search direction $\mathbf{p}^{(k)}$. How far, then, should we step in this direction? In other words, what is an appropriate step size $\alpha^{(k)}$ for the control update $\alpha^{(k)}\mathbf{p}^{(k)}$?

³ Incidentally these are also based on the idea of homotopies described in the previous chapter.

To see why this is a non-trivial question, consider the fidelity landscape in Fig. 4.2 (i.e. a maximization task). Suppose we are optimizing a seed located in the left tail of region 2 and we use local derivatives to form our step direction. A relatively large step size can be afforded that will still yield an increase in fidelity as region 2 is fairly broad and smooth. We take this step, compute new derivatives, and find that the same (relatively) large step size can be taken. We repeat this procedure many times over and climb the landscape. From this we may conclude that taking a *constant* step size is a good idea. Suppose, then, that we are optimizing a seed in the left non-zero tail of region 3. We calculate the derivatives and use the same large step size, but this time we find that the fidelity has actually decreased! We completely missed the local optimum by stepping “over” it. Re-thinking our strategy we instead use a very small step size. Similarly to before, we may find that the same small *constant* step size works great here and we quickly find the optimum. We now produce and optimize a new seed in region 2 and climb the landscape using the now-small step size. Doing this will surely locate the optima, but it now takes several orders of magnitude more iterations to get there — and calculating derivatives is an expensive affair! What is more, we are in reality blindfolded with respect to actually choosing a constant step size value in the first place as we cannot directly see the landscape and it is only by forces of luck that it will match any given local topographical situation (what about region 1?).

For these reasons it can be concluded that choosing a constant step size throughout the optimization is not the best thing to do. Too large and one might overshoot the maxima and too small one might slow the optimization to a crawl. This re-iterates that we really are dealing with *local* optimizers and different regions in optimization space can have a vastly varying local structure. To solve this problem we instead think of the locally calculated direction $\mathbf{p}^{(k)}$ as defining a particular line (or ray) in the landscape that intersects the current local point $\mathbf{u}^{(k)}$. In Fig. 4.2 the “line” is simply “left” or “right”, but the concept generalizes to

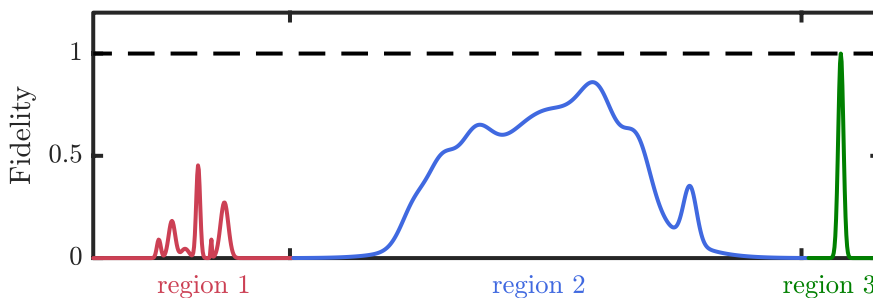


Figure 4.2: Adapted from the qualifying exam progress report. Control landscapes may exhibit different topographical structure in input space. This makes choosing an appropriate step size important and can be achieved through line searching.

higher dimensions. *Line searching* [214] is the art of choosing a proper step size $\alpha^{(k)}$ such that the cost function (minimization) is decreased an acceptable amount. That is, we can pose an auxiliary subproblem defining a one-dimensional optimization over the cost function $\tilde{J}$ as

$$\alpha^{(k)} \leftarrow \arg \min_{\alpha} \tilde{J}(\alpha) = J(\theta^{(k)} + \alpha \mathbf{p}^{(k)}), \quad \alpha \in \mathbb{R}^+, \quad (4.2)$$

where $\theta^{(k)}$ and $\mathbf{p}^{(k)}$ are fixed. In practice we are not interested in solving this sub-problem exactly but simply finding an $\alpha^{(k)}$ that gives an acceptable cost decrease. This is because chances are that the full-dimensional local minimum is not located along a given search direction as seen in Fig. 4.1 and each α probed requires evaluation of $\tilde{J}$ which entails an expensive full solution of the dynamics (N_t propagations).

This type of procedure is therefore called *inexact*⁴ line searching and there are several different sets of so-called sufficient decrease conditions (stop conditions) that can be used to qualify what “acceptable” cost decrease means. A popular set of conditions are the Wolfe conditions [214] which, roughly speaking, ensures that enough decrease has been gained (the Armijo condition) and that any further gains will likely be smaller by some margin (the curvature condition). There are also several different line search methods that defines various rules of actually optimizing for $\alpha^{(k)}$ in Eq. 4.2 such as the Moré-Thuente interpolating line search [266]. Sufficient decrease conditions and line search methods can be mixed and matched depending on application. In the QENGINE [1], the step size calculation typically uses the combination of Wolfe conditions and Moré-Thuente interpolating line search. (We have also tried out alternative combinations [267–270] but the aforementioned always outperformed.) Typically these line searches increase the average computational cost per iteration from $2N_t$ time-evolutions (assuming we only need the gradient) to approximately $5N_t - 7N_t$. The additional propagations come from evaluating the cost at different points in checking the decrease conditions. In my experience, despite more than doubling the price per iteration, the gain in step size adaptation always lowers the total number of propagations by several orders of magnitude. In cases where the full Hessian is employed, the additional effort spent line searching becomes comparatively negligible.

Effective line searches are predicated on well-scaled problems. A well-scaled problem is one where all controls (i.e. the optimization parameters) have similar scale and this can always be done by “factoring out” a scale $u_n \rightarrow u_{\text{scale}} u_n$. Typically we choose u_{scale} such that $u_n \sim \mathcal{O}(1)$ since we are dealing with numerical double floating point

⁴ Note “inexact” does not refer to the cautioned “inexact derivatives” from our previous discussions. Line searching methods are actually often derivative-free internally which in this context is boonful since we do not wish to solve the sub-problem exactly and the now-fixed line $\mathbf{p}^{(k)}$ itself was obtained by appropriate derivative-based means.

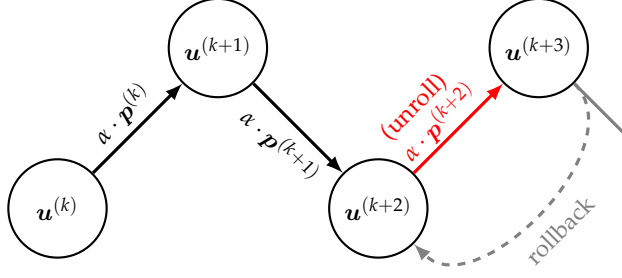


Figure 4.3: Adapted from the qualifying exam progress report. QUICK-STEP heuristic schematic. After $n_{\text{threshold}} = 2$ iterations with the same α (black edges), $n_{\text{unrolls}} = 1$ unrolls (red edge) using the same α are performed without checking decrease conditions (the step is “quickly taken”). If the step increased the cost, the solution is rolled back (grey dotted edge) and a full line search is performed (not shown). Else, check decrease conditions: perform n_{unrolls} new unrolls if they are satisfied or otherwise reset the threshold counter and continue (grey solid edge).

numbers which have a machine precision of about $\epsilon_{\text{mach}} \sim 10^{-16}$. This can also be understood as stretching or compressing the corresponding control landscape axes by a constant. Appropriate scalings become especially important if there is more than one control — if the problem is poorly-scaled the landscape would be comparatively very compressed or stretched along the separate control axes which is technically troublesome. Supposing that the problem has been well-scaled, good initial guesses for the step size in Eq. (4.2) could be $\alpha_0 = 1$ or $\alpha_0 = \alpha^{(k-1)}$. Incidentally, one attractive highlighted property of the Newton direction was that $\alpha^{(k)} \rightarrow 1$ when approaching the optimum which implies that the (otherwise potentially expensive) line search is then likely to accept the initial trial step $\alpha_0 = 1$ or can be skipped altogether.

Often we also terminate the overall optimization conditioned on the step size: if only a small step can be taken in the search direction without increasing the cost then we are likely at the optimum. This of course assumes that the local structure (and thus the new prospective search direction) does not change significantly. However this termination condition works well in practice and typical minimum threshold values are $\alpha_{\text{threshold}} = 10^{-6} - 10^{-10}$. Usually, a number of termination conditions are employed and typical choices include the optimization wall time, total number of iterations, norm of the gradient, relative decrease in cost from iteration to iteration, and so on.

Quick-Step Heuristic

Guided by the observation that constant step sizes are great as long as the local structure remains roughly the same I have developed the so-called QUICK-STEP⁵ heuristic to decrease the number of total time

⁵ Related only perhaps in spirit to the light-hearted ballroom dance of the same name.

propagations. The idea is to record subsequent step sizes found in each iteration. If the previous $n_{\text{threshold}}$ step sizes are all the same, i.e. this particular step size satisfied the sufficient decrease conditions in all these iterations, we bypass checking conditions in the next n_{unrolls} iterations and simply (greedily) use the same step size. If any of these *unrolls* produce a control that increases the cost we perform a *rollback* to undo the “bad” step and return to the previous control — from here we line search as normal and start recording the step sizes again and continue. If none of the n_{unrolls} unrolls are bad, the decrease conditions are checked in the next iteration. If they are still satisfied a new set of n_{unrolls} are performed and else a normal line search is performed. This flow is schematized in Figure 4.3 and it can be thought of in terms of balancing exploration (probing multiple candidate points in decrease conditions) vs. exploitation (greedily choosing same step size).

The heuristic incurs minimal computational overhead. The unrolling steps cost just $2N_t$ (compared to the $5N_t - 7N_t$ cited earlier) from assuming only gradient information is used in forming $\mathbf{p}$. The only costly part (N_t propagations) is associated with rolling back since we only learn that the unrolling step was bad in the process of calculating the new search direction. We essentially need to revert our actions and restore the old situation. However, if the total number of unrolls is greater than the total number of rollbacks we have still gained a massive speed up in total calculation time. Numerical investigations and experiences suggest $n_{\text{threshold}} = 3$ and $n_{\text{unrolls}} = 2 - 3$ strikes a good balance for a broad range of problems. The only caveat is that skipping the line search may produce a sub-optimal $\alpha^{(k)}$, but that possibility is already ingrained in the inexact spirit of line searching to begin with and seems to work well in practice.

This heuristic is generally applicable and agnostic towards the particular choice of decrease conditions and line search method. It could possibly be combined with momentum heuristics.

4.2 IMPOSING DISCRETIZATION — A TALE OF TWO CITIES

As already primed in previous chapters, the inevitable temporal discretization has a significant impact on the algorithmic feasibility of certain approaches. The discretization itself is always (in this thesis) taken in regular δt intervals, leading to piecewise constant controls and time evolution operators

$$t \in [0, T] \quad \rightarrow \quad \mathbf{t} = [t_1, t_2, \dots, t_{N_t}] = [0, \delta t, \dots, T], \quad (4.3a)$$

$$u \in L_2[0, T; \mathbb{R}] \rightarrow \mathbf{u} = [u_1, \dots, u_{N_t}] = [u(0), \dots, u(T)], \quad (4.3b)$$

$$\hat{\mathcal{U}}(T; 0) \approx \prod_{j=1}^{N_t-1} \hat{\mathcal{U}}_j = \hat{\mathcal{U}}_{1:N_t-1} = \hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1, \quad (4.3c)$$

where $L_2[0, T; \mathbb{R}]$ is the space of real-valued L_2 integrable functions on the interval $[0, T]$. The subscript labels from hereon will usually

refer to time indices, for example $u_n = u(t_n)$ and $\hat{U}_n = \hat{U}(u_n) = \exp(-i\hat{H}(u_n)\delta t)$ and similar for states $|\psi_n\rangle = |\psi(t_n)\rangle$ and control and drift Hamiltonians, $\hat{H}_n^c$ and $\hat{H}_n^d$, respectively. Let us remark that there is no $|\psi_{N_t+1}\rangle = |\psi(T + \delta t)\rangle$ and so there is no $\hat{U}_{N_t}$ in the propagator product. Further, a useful mnemonic is that the local forward evolution $|\psi_{n+1}\rangle = \hat{U}_n |\psi_n\rangle$ “pushes” the state *away* from index n to $n + 1$ while the backward evolution $|\psi_n\rangle = \hat{U}_n^\dagger |\psi_{n+1}\rangle$ “pulls” the state *to* index n from $n + 1$. This is mentioned only because the associated off-by-one errors are easy to make both analytically and in numerical implementation but somewhat hard and tedious to track down. Intermediate states are obtained by $|\psi_n\rangle = \hat{U}_{1:n-1} |\psi_1\rangle$ where $|\psi_1\rangle = |\psi_{\text{ini}}\rangle$. In implementation it is also worth noting that $N_t = \lfloor T/\delta t \rfloor + 1$ since T and δt are often the fixed quantities.

As in the previous chapter we consider primarily the case of a single control and only briefly touch upon the extension to multiple controls and Hamiltonians as appropriate later.

Note how Eqs. (4.3) are similar to the piecewise constant parametrization presented in Sec. 3.4 — indeed, we may think of the discretization as a type of parametrization with $\mathbf{u} = \boldsymbol{\theta}$ and $N_t = M$, i.e. the control value at each index can be individually adjusted $u_n = \theta_n$. We could similarly “bunch together” subsequent control values in intervals e.g. $M = N_t/2$ such that $\mathbf{u} = [\theta_1, \theta_1, \theta_2, \theta_2, \dots]$ and so forth. There is one important distinction, though. Everything in preceding chapter assumed that time was continuous and integrated with the full time ordered unitary propagator and δt — as used in e.g. defining parametrizations — could be of any size without loss of precision. Now upon discretization the size of δt matters as it sets both the resolution for our simulations (in the sense that $M \leq N_t$) and the precision of the time-evolution product string approximation in Eq. (4.3c). In much a similar way, the dynamics induced by the individual propagators $\hat{U}_n$ are often subjected to a further approximation that also carries with it an error typically related to some power of δt .

Apart from these fairly obvious discretization consequences there are also much more subtle and perhaps surprising ones. This is the subject of Ref. [3] and the extension in Ch. 5 by following the discretize-then-optimize approach described below. Namely, it turns out that δt also plays a role in the feasibility of producing exact and efficiently calculable control derivatives which are success criteria for local optimization as discussed elsewhere. This realization comes about by simultaneously giving discretization, problem representation, *and* the choice of numerical propagator a first-class citizen treatment at the onset of analytical control derivative calculations. The literature standard is otherwise, to my knowledge, that at minimum the last two are implicit “afterthoughts” whereas the first may [271] or may not [264] be accounted for. It is helpful to recall in parallel our earlier discussions from Sec. 3.8 where it was argued that the efficacy of theoretical

open-loop results is tied to how well the underlying model aligns with the experimental reality. Quite analogously, within the purely theoretical open-loop setting *itself*, the numerical implementation details play the role of the “experimental reality” and our analytical derivations should therefore likewise seek to accommodate these. In particular, we find certain numerical choices of propagator and representation — a Suzuki-Trotter-type expansion and a diagonalized basis for the control Hamiltonian(s) — to be preferable out of all the standard propagators and types of representation.

We will cover two lines of derivations for the control derivatives proceeding from the two highlighted versions of Eq. (3.1) and equivalently Eq. 3.3, individually characterized by the order in which one optimizes⁶ and discretizes:

Optimize-then-discretize [1] — we follow the precedent set in Ch. 3 and calculate the functional Gâteaux variations using Eq. (3.5) and from these identify the (continuous) derivatives $\nabla_{u(t)} J[u^{(k)}](t)$ and $\nabla_u^2 J[u^{(k)}](t)$ that enter into forming the search directions $p^{(k)}(t)$ and $\mathbf{p}^{(k)}$ in Eqs. (3.9), (3.17). This leads to a system of three coupled differential equations at which point we impose the discretization of Eqs. (4.3) and attain finite-dimensional versions e.g. $\nabla_{u(t)} J[u](t) \rightarrow \overline{\nabla_u J}(\mathbf{u})$ (overlines in this chapter do not refer to ensemble costs as in Sec. (3.7)).

Discretize-then-optimize [3] — we immediately impose the discretization Eqs. (4.3) and calculate ordinary vector derivatives $\nabla_u J(\mathbf{u})$ and $\nabla_u^2 J(\mathbf{u})$ for the search directions $\mathbf{p}^{(k)}$. This approach more naturally encompasses the separate treatment for individual time-evolution operator approximants. The ensuing derivatives for the considered suite of approximants are very different and the control-diagonal Trotter-type is found to have the best properties. Regardless, we are lead to a similar set of equations as in the optimize-then-discretize approach. Generally, however, $\nabla_u J(\mathbf{u}) \neq \overline{\nabla_u J}(\mathbf{u})$, with one special exception where an equality is found — namely for the very same control-diagonal Trotter-type approximant mentioned above.

For our purposes the ultimately best of these two proposals is the latter because it takes into account the numerical implementation details. The earlier due diligence in establishing control landscapes will also come into play in this understanding and interpretation. What is more, there is a somewhat steep buy-in for following the optimize-then-discretize path since we need to invoke comparatively involved machinery from the calculus of variations. This results in

⁶ “Optimize” in this context does not refer to process of actually traversing the control landscape but rather the act of analytically calculating and zeroing the variation — which leads to working derivatives that are then used to “optimize” solutions in the usual sense.

a few relatively arcane mathematical objects and subtleties that one must either accept at face value or devote a certain length of time to appreciate more deeply. Following the discretize-then-optimize path, on the other hand, requires only ordinary partial derivatives and is in this particular sense much more approachable. In fact, the same mathematical objects develop organically in now not-so-arcanic terms. It is important to emphasize that the results hereof require only standard numerical devices past the initial analytical derivations.

4.3 OPTIMIZE-THEN-DISCRETIZE

Here we derive the the so-called optimality equations pertaining to the state transfer of BEC. Emphasis is put on thoroughness of the steps and relevant equations for the derivations are re-stated for convenience where applicable so as to be self-contained.

We consider a BEC meanfield described by the Gross-Pitaevskii Hamiltonian

$$\hat{H}_{\text{gp}} = \hat{T} + \hat{V} + g|\psi|^2 = \hat{H}_0 + g|\psi|^2, \quad (4.4)$$

where $\hat{H}_0 = \hat{T} + \hat{V}(u(t))$ is the interaction-free part and the last term is the condensate self-interaction with strength g . The following derivation is valid for any dimensionality and physical system governed only by some $\hat{H}_0$ (not only those described by real-space wave functions), corresponding to setting $g = 0$ throughout.

The time evolution of the condensate is governed by the time-dependent Gross-Pitaevskii equation

$$i\partial_t |\psi\rangle = (\hat{H}_0 + g|\psi|^2) |\psi\rangle, \quad (4.5)$$

where $\hbar = 1$. In the state transfer context, our goal is to find a control function $u(t)$ that realizes an $\hat{H}_0(u(t))$ such that the transfer $|\psi_{\text{ini}}\rangle \rightarrow |\psi_{\text{tgt}}\rangle$ occurs with high fidelity, $F = |\langle \psi_{\text{tgt}} | \psi(T) \rangle|^2$, where T is the process duration. This task can be posed as a minimization of the total cost functional J

$$J[\psi, \chi, u] = J_F[\psi] + J_{\text{gp}}[\psi, \chi, u] + J_\gamma[u], \quad (4.6)$$

Note that the inputs to J are *functions*, i.e. $J[\psi(t), \chi(t), u(t)]$, but the dependence on time is often omitted for brevity. The total cost⁷ is composed of specialized functionals from Sec. 3.6,

$$J_F[\psi] = \frac{1}{2} (1 - F) = \frac{1}{2} (1 - |\langle \psi_{\text{tgt}} | \psi(T) \rangle|^2), \quad (4.7a)$$

$$J_{\text{gp}}[\psi, \chi, u] = \Re \int_0^T \langle \chi | [i\partial_t - \hat{H}_0(u) - g|\psi|^2] |\psi\rangle dt, \quad (4.7b)$$

$$J_\gamma[u] = \frac{\gamma}{2} \int_0^T \dot{u}^2 dt. \quad (4.7c)$$

⁷ We could also have include even more terms here, but choose only these for later comparison to other works.

The fidelity cost quantifies how well the state was transferred and the derivative regularization cost shifts preference towards smoother controls depending on the value of γ . The second cost ensures that the equation of motion for ψ is obeyed at all time. Relatedly, the fidelity cost is here considered only a functional in ψ , and not $u(t)$, because the dynamics is accounted for through the second term. The “price” we pay for this formulation is that we must also determine and solve the equation of motion for the auxiliary state $|\chi(t)\rangle$ which is also called the Lagrange multiplier, co-state, or adjoint state.

To derive the (first-order) optimality equations for the problem at hand, we need to calculate the Gâteaux derivative (or variation) of J with respect to all the input functions and require these to vanish since this is what characterizes an optimal point (ψ^*, χ^*, u^*) [216]. We begin with the general definition of the Gâteaux derivative and the associated notion of gradients.

4.3.1 The Gâteaux Derivative

The Gâteaux derivative is the generalized directional derivative. Suppose we have a mapping $f : X \rightarrow Y$, where X and Y are normed vector spaces. Then the Gâteaux derivative at $g \in X$ in the $\delta h \in X$ direction is

$$D_{\delta h}f(g) = \lim_{\epsilon \rightarrow 0} \frac{f(g + \epsilon \delta h) - f(g)}{\epsilon} = \frac{d}{d\epsilon} \left(f(g + \epsilon \delta h) \right)_{\epsilon=0}. \quad (4.8)$$

The first equality is primarily useful for a numerical finite-differencing implementation, and the second equality is primarily useful for analytical derivation.

We define the *gradient* of f at g as the unique element $\nabla_X f \in X$ in terms of the directional derivative

$$D_{\delta h}f(g) = \langle \nabla_X f(g), \delta h \rangle_X, \quad (4.9)$$

where $\langle \cdot, \cdot \rangle_X$ is the norm defined on X . This relation is very useful, since it allows us to implicitly calculate $\nabla_X f$ by comparing the result of Eq. (4.8). It is emphasized that $\nabla_X f \in X$ resides in the input space and $D_{\delta h}f \in Y$ in the output space. The second variation and associated Hessian is similarly defined, but we do not consider this further here.

Example: $f : \mathbb{R}^N \rightarrow \mathbb{R}$, i.e. $f(\mathbf{x})$ is a function accepting a vector of length N and outputting a number, e.g. $f(\mathbf{x}) = f(x_1, x_2) = -(x_1^2 + x_2^2)$.

The Gâteaux derivative of f in the $\delta\mathbf{x} = (\delta x_1, \delta x_2)^T$ direction at the point $\mathbf{x} = (x_1, x_2)^T$ is

$$\begin{aligned} D_{\delta\mathbf{x}}f(\mathbf{x}) &= -\frac{d}{d\epsilon}((x_1 + \epsilon\delta x_1)^2 + (x_2 + \epsilon\delta x_2)^2)|_{\epsilon=0} \\ &= -(2(x_1 + \epsilon\delta x_1)\delta x_1 + 2(x_2 + \epsilon\delta x_2)\delta x_2)|_{\epsilon=0} \\ &= -2(x_1\delta x_1 + x_2\delta x_2) = (-2\mathbf{x}) \cdot \delta\mathbf{x}. \end{aligned} \quad (4.10)$$

Let us consider the standard inner product for $X = \mathbb{R}^N$,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_i x_i y_i. \quad (4.11)$$

Using the definition of the gradient and the above calculation leads to

$$D_{\delta\mathbf{x}}f(\mathbf{x}) \stackrel{(4.9)}{=} \langle \nabla f, \delta\mathbf{x} \rangle \stackrel{(4.11)}{=} \nabla f \cdot \delta\mathbf{x} \stackrel{(4.10)}{=} (-2\mathbf{x}) \cdot \delta\mathbf{x}. \quad (4.12)$$

We can therefore conclude that $\nabla f = -2\mathbf{x}$ but only on the grounds that $\delta\mathbf{x}$ is an arbitrary variation. This is consistent with the result from the more familiar definition of the gradient in multivariate calculus $\nabla f = (\partial_{x_1}f, \partial_{x_2}f)^T$.

We could also, however, have chosen any other inner product for $\mathbb{R}^N$ which are all on the general form,

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \mathbf{x}^T \mathbf{A} \mathbf{y} = \mathbf{x}^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mathbf{y}, \quad (4.13)$$

where $\mathbf{A}$ is symmetric and positive-definite (strictly positive eigenvalues and thus invertible). After similar comparative lines of calculations the gradient *with respect to the inner product* defined by $\mathbf{A}$ can be written

$$\nabla_A f = \mathbf{A}^{-1}(2\mathbf{x}) = \begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} (2\mathbf{x}), \quad (4.14)$$

which reduces to the above result when $\mathbf{A}$ is the identity ($a = c = 1$ and $b = 0$). The point is that $D_{\delta\mathbf{x}}f(\mathbf{x})$ is always the same, but the the gradient depends on the choice of inner product. $\square$

The enormous advantage of the Gâteaux formulation is that it applies to any mapping between normed vector spaces X and Y . In this sense, it provides a formal procedure for calculating derivatives with respect to e.g. scalars, vectors, matrices, functions, etc., on the same footing. In our particular case we have the cost functional $J[\psi(t), \chi(t), u(t)]$ where the inputs to J are functions of t defined on the interval $[0, T]$. The natural inner product is the L_2 inner product

$$\langle f(t), g(t) \rangle_{L_2} = \int_0^T f(t)g(t)dt. \quad (4.15)$$

Another choice of inner product is the H_1 inner product [264]

$$\langle f(t), g(t) \rangle_{H_1} = \int_0^T \dot{f}(t) \dot{g}(t) dt. \quad (4.16)$$

As in the example above we now have to be careful when addressing gradients: according to Eq. (4.9) we cannot talk about *the* gradient, but rather the gradient *with respect to the chosen inner product* defined on X . Specifically we will distinguish in this section between the L_2 gradient ($\nabla_{L_2} J$), and the H_1 gradient ($\nabla_{H_1} J$). It is also implicitly assumed that we are talking about the gradient with respect to $u(t)$.

We now show how $\nabla_{H_1} J$ can be calculated from $\nabla_{L_2} J$. Noting that Eq. (4.9) holds equal for all choices of X :

$$\begin{aligned} \langle \nabla_{H_1} J(t), \delta u(t) \rangle_{H_1} &= \int_0^T \nabla_{H_1} \dot{J}(t) \delta \dot{u}(t) dt = \int_0^T (-\nabla_{H_1} \ddot{J}(t)) \delta u(t) dt \\ &= \langle \nabla_{L_2} J(t), \delta u(t) \rangle_{L_2} = \int_0^T \nabla_{L_2} J(t) \delta u(t) dt, \end{aligned} \quad (4.17)$$

where we used integration by parts in the second equality and assumed the variation δu is zero on the boundaries (a basic assumption which will continue in the next section). We can thus conclude by comparison

$$\nabla_{H_1} \ddot{J}(t) = -\nabla_{L_2} J(t), \quad (4.18)$$

which is a Poisson equation in t . In other words, if we can calculate the L_2 gradient we can also find the H_1 gradient. The L_2 gradient is the program for the next section.

4.3.2 Optimality Equations

We now derive the first-order optimality equations by calculating and zeroing the three Gâteaux variations

$$D_{\delta\psi} J = D_{\delta\chi} J = D_{\delta u} J = 0, \quad \text{for all } \delta\psi, \delta\chi, \delta u, \quad (4.19a)$$

under the constraints that

$$\delta\psi(0) = 0, \quad \text{and} \quad \delta u(0) = \delta u(T) = 0. \quad (4.19b)$$

Note that apart from the constraints, the variations are *completely arbitrary*. This fact is extensively utilized, as in the example, to reach certain conclusions in the following.

Variation of χ

Setting the definition of the Gâteaux variation to zero we obtain

$$\begin{aligned}
 0 = D_{\delta\chi}J &= \frac{d}{d\epsilon} (J_F[\psi] + J_{\text{gp}}[\psi, \chi + \epsilon\delta\chi, u] + J_\gamma[u])_{\epsilon=0} \\
 &= \left(\Re \int_0^T \frac{d}{d\epsilon} (\langle \chi | + \epsilon \langle \delta\chi |) [i\partial_t - \hat{H}_0(u) - g|\psi|^2] |\psi\rangle dt \right)_{\epsilon=0} \\
 &= \Re \int_0^T \langle \delta\chi | [i\partial_t - \hat{H}_0(u) - g|\psi|^2] |\psi\rangle dt. \tag{4.20}
 \end{aligned}$$

Invoking the fact that the variation $\delta\chi$ is arbitrary leads us to conclude that

$$i\partial_t |\psi\rangle = (\hat{H}_0 + g|\psi|^2) |\psi\rangle, \tag{4.21}$$

which is exactly the time-dependent Gross-Pitaevskii equation.

Variation of ψ

This variation takes a bit of work. We start with

$$0 = D_{\delta\psi}J = \frac{d}{d\epsilon} \left(J_F[\psi + \epsilon\delta\psi] + J_{\text{gp}}[\psi + \epsilon\delta\psi, \chi, u] + J_\gamma[u] \right)_{\epsilon=0}. \tag{4.22}$$

Only the two first terms involve ψ . We will use the following identities for a complex function $\kappa = \kappa(\epsilon)$,

$$\kappa + \kappa^* = 2\Re(\kappa) = 2\Re(\kappa^*), \tag{4.23a}$$

$$\frac{d}{d\epsilon} |\kappa|^2 = \frac{d}{d\epsilon} (\kappa^* \kappa) = \frac{d\kappa^*}{d\epsilon} \kappa + \kappa^* \frac{d\kappa}{d\epsilon} = 2\Re \left(\kappa^* \frac{d\kappa}{d\epsilon} \right), \tag{4.23b}$$

$$\Re(\Re(\kappa)) = \Re(\kappa). \tag{4.23c}$$

We calculate the two terms in turn. First,

$$\begin{aligned}
 D_{\delta\psi}J_F &= \frac{d}{d\epsilon} \left(J_F[\psi + \epsilon\delta\psi] \right)_{\epsilon=0} \\
 &= \frac{d}{d\epsilon} \left(\frac{1}{2} (1 - |\langle \psi_{\text{tgt}} | (\psi(T)\rangle + \epsilon \langle \delta\psi(T)\rangle)|^2) \right)_{\epsilon=0} \\
 &= -\Re \langle \psi(T) | \psi_{\text{tgt}} \rangle \langle \psi_{\text{tgt}} | \delta\psi(T) \rangle, \tag{4.24}
 \end{aligned}$$

where we used Eq. (4.23b) with $\kappa(\epsilon) = \langle \psi_{\text{tgt}} | \psi(T)\rangle + \epsilon \langle \psi_{\text{tgt}} | \delta\psi(T)\rangle$. Next we calculate

$$\begin{aligned}
 D_{\delta\psi}J_{\text{gp}} &= \frac{d}{d\epsilon} \left(J_{\text{gp}}[\psi + \epsilon\delta\psi, \chi, u] \right)_{\epsilon=0} \\
 &= \frac{d}{d\epsilon} \left(\Re \int_0^T \langle \chi | \left[i\partial_t - \hat{H}_0(u) - g|\psi + \epsilon\delta\psi|^2 \right] (|\psi\rangle + \epsilon |\delta\psi\rangle) dt \right)_{\epsilon=0}. \tag{4.25}
 \end{aligned}$$

Expanding the last term in the square bracket onto the spatial basis (for the moment) and taking the derivative gives

$$\begin{aligned} \left. \frac{d}{d\epsilon} \left(\chi^* |\psi + \epsilon \delta\psi|^2 (\psi + \epsilon \delta\psi) \right) \right|_{\epsilon=0} &= \chi^* \left(\delta\psi^* (\psi + \epsilon \delta\psi) (\psi + \epsilon \delta\psi) \right. \\ &\quad \left. + (\psi + \epsilon \delta\psi)^* \delta\psi (\psi + \epsilon \delta\psi) \right. \\ &\quad \left. + (\psi + \epsilon \delta\psi)^* (\psi + \epsilon \delta\psi) \delta\psi \right) \Big|_{\epsilon=0} \\ &= \chi^* \psi^2 \delta\psi^* + 2\chi^* |\psi|^2 \delta\psi. \end{aligned} \quad (4.26)$$

Our goal is to isolate everything as a product with the variation $\delta\psi$, so we need to rewrite $\delta\psi^*$ in terms of $\delta\psi$. Rewriting Eq. (4.23a) into e.g. $\kappa = 2\Re(\kappa^*) - \kappa^*$ with $\kappa = \chi^* \psi^2 \delta\psi^*$:

$$\chi^* \psi^2 \delta\psi^* = 2\Re(\chi \psi^{*2} \delta\psi) - \chi \psi^{*2} \delta\psi. \quad (4.27)$$

Pulling in the real part from outside the integral and using Eq. (4.23c) gives

$$\begin{aligned} \Re(\chi^* \psi^2 \delta\psi^*) &= \Re(2\Re(\chi \psi^{*2} \delta\psi) - \chi \psi^{*2} \delta\psi) = 2\Re(\chi \psi^{*2} \delta\psi) - \Re(\chi \psi^{*2} \delta\psi) \\ &= \Re(\chi \psi^{*2} \delta\psi). \end{aligned} \quad (4.28)$$

Inserting the result in bra-ket form yields

$$= \Re \int_0^T \left(\langle \chi | [i\partial_t - \hat{H}_0(u) - 2g|\psi|^2] - \langle \chi^* | g\psi^{*2} \right) |\delta\psi\rangle dt, \quad (4.29)$$

where $\langle \chi^* |$ signifies that χ is to be complex conjugated when expanding on the spatial basis. The conjugate of the bra retains the “inner” complex conjugation, $(\langle \chi^* |)^\dagger = |\chi^*\rangle$.

The final step is to move the time derivative in the first term from $|\delta\psi\rangle$ to $\langle \chi |$ using integration by parts:

$$\int_0^T i \langle \chi | \delta\psi \rangle dt = i \langle \chi(T) | \delta\psi(T) \rangle - i \langle \chi(0) | \delta\psi(0) \rangle - \int_0^T i \langle \dot{\chi} | \delta\psi \rangle dt. \quad (4.30)$$

The second term above is zero since we enforce the constraint $\delta\psi(0) = 0$. We can then combine the two Gâteaux derivatives to get

$$\begin{aligned} 0 &= D_{\delta\psi} J = \Re \left([-\langle \psi(T) | \psi_{\text{tgt}} \rangle \langle \psi_{\text{tgt}} | + i \langle \chi(T) |] |\delta\psi(T)\rangle \right) \\ &\quad - \Re \int_0^T [i \langle \dot{\chi} | + \langle \chi | (\hat{H}_0(u) + 2g|\psi|^2) + \langle \chi^* | g\psi^{*2}] |\delta\psi\rangle dt. \end{aligned} \quad (4.31)$$

Now we are ready to invoke that $\delta\psi$ is an arbitrary variation. For $t = T$ we obtain from the first line square bracket above

$$-\langle \psi(T) | \psi_{\text{tgt}} \rangle \langle \psi_{\text{tgt}} | + i \langle \chi(T) | = 0 \Rightarrow |\chi(T)\rangle = i \langle \psi_{\text{tgt}} | \psi(T) \rangle |\psi_{\text{tgt}}\rangle. \quad (4.32)$$

This is the initial condition for the Lagrange multiplier. For $0 \leq t \leq T$ we obtain from the second line square bracket

$$\begin{aligned} i \langle \dot{\chi} | + \langle \chi | (\hat{H}_0(u) + 2g|\psi|^2) + \langle \chi^* | g\psi^{*2} &= 0, \\ \Rightarrow i \langle \dot{\chi} | &= (\hat{H}_0(u) + 2g|\psi|^2) |\chi\rangle + g\psi^2 |\chi^*\rangle. \end{aligned} \quad (4.33)$$

This is the equation of motion for the Lagrange multiplier. Note that only in the case of $g = 0$ does the Lagrange multiplier obey the usual Schrödinger equation — then the interim projection into the spatial basis is unnecessary and the results hold for general systems.

Variation of u

We turn to the variation of u with

$$0 = D_{\delta u} J = \frac{d}{d\epsilon} \left(J_F[\psi] + J_{\text{gp}}[\psi, \chi, u + \epsilon \delta u] + J_\gamma[u + \epsilon \delta u] \right)_{\epsilon=0}. \quad (4.34)$$

Only the two last terms depend on ϵ . Calculating them in turn gives

$$\begin{aligned} D_{\delta u} J_{\text{gp}} &= \frac{d}{d\epsilon} \left(\Re \int_0^T \langle \chi | [i\partial_t - \hat{H}_0(u + \epsilon \delta u) - g|\psi|^2] |\psi\rangle dt \right)_{\epsilon=0} \\ &= \left(\Re \int_0^T \langle \chi | \left[0 - \frac{d\hat{H}_0(u + \epsilon \delta u)}{d\epsilon} - 0 \right] |\psi\rangle dt \right)_{\epsilon=0} \\ &= \left(-\Re \int_0^T \langle \chi | \left[\frac{d\hat{H}_0(u + \epsilon \delta u)}{d(u + \epsilon \delta u)} \frac{d(u + \epsilon \delta u)}{d\epsilon} \right] |\psi\rangle dt \right)_{\epsilon=0} \\ &= -\Re \int_0^T \langle \chi | \frac{d\hat{H}_0(u)}{du} |\psi\rangle \delta u dt. \end{aligned} \quad (4.35)$$

Moving on to the J_γ derivative we have

$$\begin{aligned} D_{\delta u} J_\gamma &= \frac{d}{d\epsilon} \left(\frac{\gamma}{2} \int_0^T \left(\frac{d}{dt}(u + \epsilon \delta u) \right)^2 dt \right)_{\epsilon=0} \\ &= \left(\frac{\gamma}{2} \int_0^T 2(\dot{u} + \epsilon \dot{\delta u}) \frac{d}{d\epsilon}(\dot{u} + \epsilon \dot{\delta u}) dt \right)_{\epsilon=0} \\ &= \gamma \int_0^T \dot{u} \delta u dt = \gamma \left[\dot{u}(T) \delta u(T) - \dot{u}(0) \delta u(0) - \int_0^T \ddot{u} \delta u \right] \\ &= -\gamma \int_0^T \ddot{u} \delta u dt. \end{aligned} \quad (4.36)$$

where we used integration by parts and that $\delta u(0) = \delta u(T) = 0$ as per our constraints. Adding the two terms together yields

$$0 = D_{\delta u} J = \int_0^T \left(-\Re \langle \chi | \frac{d\hat{H}_0(u)}{du} |\psi\rangle - \gamma \ddot{u} \right) \delta u \quad (4.37)$$

Invoking the fact that δu is arbitrary and reading off the gradient by comparing to $D_{\delta u} J = \langle \nabla_{L_2} J, \delta u \rangle_{L_2}$ we obtain the final optimality equation

$$0 = \nabla_{L_2} J = -\Re \langle \chi | \frac{d\hat{H}_0(u)}{du} |\psi\rangle - \gamma \ddot{u}.$$

4.3.3 Summary of Optimality Equations

By taking the first Gâteaux variation of J with respect to all the input functions, i.e. the triplet $\{\psi(t), \chi(t), u(t)\}$, we obtained the following first order optimality system:

$$i|\dot{\psi}\rangle = (\hat{H}_0(u) + g|\psi|^2)|\psi\rangle, \quad (4.38a)$$

$$i|\dot{\chi}\rangle = (\hat{H}_0(u) + 2g|\psi|^2)|\chi\rangle + g\psi^2|\chi^*\rangle, \quad (4.38b)$$

$$0 = \nabla_{L_2} J = -\Re\langle\chi|\frac{d\hat{H}_0(u)}{du}|\psi\rangle - \gamma\ddot{u}, \quad (4.38c)$$

with initial conditions for the states and control

$$|\psi(0)\rangle = |\psi_{\text{ini}}\rangle, \quad (4.38d)$$

$$|\chi(T)\rangle = io|\psi_{\text{tgt}}\rangle, \quad (4.38e)$$

$$u(0) = u_0, u(T) = u_T. \quad (4.38f)$$

Here we defined the transfer amplitude $o = \langle\psi_{\text{tgt}}|\psi(T)\rangle$. This optimality system is quite satisfying in the sense that it tells us that $|\psi\rangle$ must follow the relevant quantum dynamics and that the gradient with respect to u must vanish, e.g. $\nabla_{L_2} J = 0$ (and similar for the H_1 gradient obtained by Eq. (4.18)). Less obvious is the immediate physical interpretation of $|\chi\rangle$ but we can merely think of it as a mathematical tool.

Actually, the L_2 gradient may not vanish at the boundaries so we must always artificially enforce $\nabla_{L_2} J(0) = \nabla_{L_2} J(T) = 0$. In solving the Poisson equation for the H^1 gradient, however, we may conveniently choose Dirichlet boundary conditions $\nabla_{H^1} J(0) = \nabla_{H^1} J(T) = 0$ directly⁸ so the end points always vanish by construction. In a number of papers (e.g. [264] and references thereto) the H_1 gradient is found to be preferable due to its inherent smoothness and overall performance. With respect to the latter point, that trend has been elusive in my studies and the L_2 version has seemed to always outperform and can be smoothed out by other means. This is therefore of no real further concern especially since we nowadays internally follow the discretize-then-optimize approach which is e.g. unhampered by the potentially non-vanishing gradient at the boundaries. From hereon we will occasionally drop the norm identifier subscripts and in those cases assume we are referring to the $X = L_2$.

4.3.4 Issues Arising and Discretizing the Gradient

There are at least three problems with the stated optimality equations 4.38.

The first and most immediate problem is that they are not analytically soluble outside the most simple limiting cases [222, 272] — their

⁸ "dirichtly", as it were.

coupling and interplay is simply too hard to integrate by hand. This is what prompts a numerical iterative treatment⁹.

The second problem is that they are only necessary, but not sufficient, conditions for qualifying a minimum of J . Since the Gâteaux variations are generalized directional derivatives, zeroing them is an effort in finding stationary points. As such, the optimality system is valid for *all* stationary points, i.e. also for solutions that are local maxima. Sufficient conditions are related to the local curvature and thus the local control Hessian [215]. (The control Hessian is defined by the second variation of u and is straightforward to calculate but omit it here for brevity). As it turns out, fortunately, concerns related to this type of sufficient condition is not really a problem¹⁰ in the iterative schemes because the steps themselves are constructed such that the sequence of controls decreases the cost. As a matter of fact, with an iterative scheme for u in mind and now equipped with the gradient Eq. (4.38c) we can return to thinking about the cost as just a functional in the control, $J[\psi, \chi, u] \rightarrow J[u(t)]$, so long as we solemnly promise to solve for $|\psi\rangle$ and $|\chi\rangle$ using the equations of motion Eqs. (4.38a)–(4.38b) (this is also sometimes called the “reduced” cost [216]).

The third and more relevant issue is that the optimality equations are not yet discretized in time. We remark intermittently that there is a certain gauge freedom in choosing both the overall sign, scaling, and in using the real $\Re$ or imaginary $\Im$ part in the constraint cost Eq. (4.7b) — compare e.g. Ref. [264] and Ref. [273]. The convention of the former was chosen for the above derivations. This gauge freedom is allowed due to $|\chi\rangle$ being an auxiliary, purely mathematical object that will adapt accordingly to the chosen convention. In anticipation of comparative developments later in the exposition, we shall now effectively rescale $\langle\chi| \rightarrow -io^* \langle\chi|$ by factoring out $io = i \langle\psi_{\text{tgt}}|\psi(T)\rangle$ from its initial condition Eq. (4.38e) such that now $|\chi(T)\rangle = |\psi_{\text{tgt}}\rangle$. After this change in convention and upon piecewise constant δt -discretization,

⁹ Although written as three coupled differential equations and thus suggesting explicit ode integration, it is more preferable to use employ a unitary propagator approach for the dynamics as described later in Sec. 4.5.2.

¹⁰ Unless of course the starting point is already a local maximum in which case the gradient vanishes and thus also the step itself. It is nevertheless an exceedingly exceptional occurrence to start in any optima.

i.e. $u(t) = u_n$ over the interval $[t_n, t_n + \delta t]$, we find the gradient elements

$$\begin{aligned}
 \nabla J(t) &\rightarrow (\overline{\nabla u J})_n = \int_{t_n}^{t_n + \delta t} \nabla J(t) dt \\
 &= \int_{t_n}^{t_n + \delta t} \Re \left(io^* \left\langle \chi \left| \frac{d\hat{H}_0(u)}{du} \right| \psi \right\rangle \right) dt \\
 &= \Re \left(io^* \left\langle \chi \left| \frac{d\hat{H}_0(u)}{du} \right| \psi \right\rangle \right) \Big|_{t=t_n} \int_{t_n}^{t_n + \delta t} dt \\
 &= \Re \left(io^* \left\langle \chi_n \left| \frac{d\hat{H}_0(u_n)}{du} \right| \psi_n \right\rangle \right) \delta t, \quad (4.39)
 \end{aligned}$$

for all time indices $n = 2 : N_t - 1$ (see below) where we took $\gamma = 0$ for brevity. Note that $|\chi_{N_t}\rangle = |\psi_{\text{tgt}}\rangle$.

BOUNDARY BEHAVIOR — We record the boundary behavior of the derivatives for later comparisons. The (discrete) gradient expression is valid for $n = 2 : N_t - 1$ because we assumed $\delta u(0) = \delta u(T) = 0$. We could subsume this explicitly by multiplying the elements with $\bar{E}_n = \delta_{n,1} + \delta_{n,N_t}$, where $\delta_{n,m} = 1 - \delta_{n,m}$ is the “anti”-Kronecker delta.

4.3.5 Constraint Cost Gradients

Adding more costs to the total functional requires us to also include the associated (discretized) gradients. These are straightforward to calculate with the same procedure as above. We have already seen the derivative regularization and listing a few more of them we have

$$(\overline{\nabla J_\alpha})_n = \alpha u_n, \quad (4.40a)$$

$$(\overline{\nabla J_\gamma})_n = -\gamma \ddot{u}_n, \quad (4.40b)$$

$$(\overline{\nabla J_\sigma})_n = \sigma \cdot \left(\Theta(u_{\min} - u_n)(u_n - u_{\min}) + \Theta(u_n - u_{\max})(u_n - u_{\max}) \right). \quad (4.40c)$$

Note that in writing $\ddot{u}$ we have still yet to apply an appropriate finite differencing stencil scheme. We will not carry this out in detail here because we will depart from this line of approach momentarily.

4.3.6 Control Parametrization

We can extend our results to also cover the control parametrization $u = u(t; \theta)$ as in Sec. 3.5 where $\theta \in \mathbb{R}^M$. The new optimality system is then formed by calculating the Gâteaux variations now with respect to the triple $\{\psi(t), \chi(t), \theta\}$

$$D_{\delta\psi}J = D_{\delta\chi}J = D_{\delta\theta}J = 0. \quad (4.41)$$

The first two produce exactly the same equations of motion as before and when we solemnly promise to adhere to these we effectively

have $J = J(\theta)$. Choosing the inner product to be the usual vector dot product for $X = \mathbb{R}^M$ (as in the example Eq. (4.11)), the variation in θ reads (using e.g. a variational chain rule)

$$\begin{aligned} D_{\delta\theta}J &= \langle \nabla_{\theta}J, \delta\theta \rangle = \frac{d}{d\epsilon} (J[u(t; \theta + \epsilon\delta\theta)])_{\epsilon=0} \\ &= \left(\int_0^T \left\{ -\Re \left\langle \chi \left| \frac{d\hat{H}_0(u)}{du} \right| \psi \right\rangle \right\} \nabla_{\theta}u \, dt \right) \cdot \delta\theta \\ &= \left\langle \left(\int_0^T \left\{ -\Re \left\langle \chi \left| \frac{d\hat{H}_0(u)}{du} \right| \psi \right\rangle \right\} \nabla_{\theta}u \, dt \right), \delta\theta \right\rangle, \end{aligned} \quad (4.42)$$

where we for brevity both omitted the regularization term ($\gamma = 0$) and used the “old” $\langle \chi |$ convention. Note the integrand is a vector (*qua* $\nabla_{\theta}u$) and the (vector) integral is dotted with the time-independent vector variation $\delta\theta$. By comparative inspection of the first and last equality and using that $\delta\theta$ is an arbitrary variation we conclude

$$\nabla_{\theta}J = \int_0^T \left\{ -\Re \left\langle \chi \left| \frac{d\hat{H}_0(u)}{du} \right| \psi \right\rangle \right\} \nabla_{\theta}u \, dt = \int_0^T \nabla_{L^2}J(t) \nabla_{\theta}u(t) \, dt. \quad (4.43)$$

This means that we must always, no matter what convention or parametrization, calculate the gradient $\nabla_{L^2}J(t)$ and then integrate it alongside the vector derivative $\nabla_{\theta}u(t; \theta)$.

In the time-discretized setting we of course need to apply an appropriate numerical integration quadrature for the individual elements, for example the simplest piecewise constant with uniform sampling ($\int_0^T \dots dt \rightarrow \sum_{i=1}^{N_t} \dots \delta t$),

$$(\nabla_{\theta}J)_n \rightarrow (\overline{\nabla_{\theta}J})_n = \sum_{i=1}^{N_t} (\overline{\nabla_{L^2}J})_i \left(\frac{\partial u_i}{\partial \theta_n} \right) \quad (4.44)$$

where it is understood that $u_i = u(t_i; \theta)$. Note that δt is subsumed in the discretized L_2 gradient. The only real requirement here is, unsurprisingly, that u is composed of differentiable functions with respect to θ .

Examples

As a sanity check we may consider a parametrization (cf. Sec. 3.5) into piecewise constant functions $u = \sum_n f_n^{\text{uniform}}(t; \theta_n) = \sum_n I(t \in r_n) \theta_n$ with full resolution ($M = N_t$) such that each time index n has an individually adjustable control parameter θ_n . Then

$$(\overline{\nabla_{\theta}J})_n = \sum_{i=1}^{N_t} (\overline{\nabla_{L^2}J})_i I(t_i \in r_n) = (\overline{\nabla_{L^2}J})_n, \quad (4.45)$$

which correctly yields the expected result Eq. (4.39) up to a relabeling of the optimization variable.

We can also evaluate the gradient in a CRAB-like basis expansion for example $u = u_{\text{ref}} + \Gamma(t) \sum_n f_n^{\text{Fourier}}(t; \{c_n^r, \omega_n^r, \phi_n^r\}_{r=c,s})$ as described in Sec. 3.5 (here we factored out the common shape function Γ). Focusing just on the expansion coefficients we attain the form (compare Ref. [165])

$$(\nabla_c J)_n = \sum_{i=1}^{N_t} (\nabla_{J_{L_2}})_i \Gamma_i \frac{\partial f_{n,i}^{\text{Fourier}}}{\partial c_n}, \quad (4.46)$$

and similar for the frequency and phase parameters if so desired.

From these examples it is easy to see how any appropriate analytical parametrization can, indeed, be subjected to derivative-based updates¹¹.

4.4 DISCRETIZE-THEN-OPTIMIZE

In the discretize-then-optimize approach we immediately apply the discretization Eq. (4.3) and write out the time-evolution in terms of propagators,

$$J(\mathbf{u}) = \frac{1}{2} (1 - |\langle \psi_{\text{tgt}} | \hat{\mathcal{U}}_{1:N_t-1} | \psi_{\text{ini}} \rangle|^2). \quad (4.47)$$

For simplicity we consider here $J = J_F$ as the total cost, thus dropping the subscript momentarily, and defer the constraint costs until later. What is notable here compared to the optimize-then-discretize approach is not only that we have already discretized, but we have also incorporated the quantum dynamics in the fidelity term instead of an auxiliary cost J_{gp} .

The Gâteaux variational procedure that embodied the last section is very flexible and we could also equally well apply it here to obtain the derivatives. The form of Eq. (4.47), however, is a regular vector function in $\mathbf{u}$ and thus unencumbered from the functional structure that prompted the Gâteaux formulation. Instead, we can with leisure just consider ordinary partial derivatives with respect to each control element $(\mathbf{u})_n = u_n$. Note that this is now just a special case $M = N_t$ of the piecewise constant parametrization as in the example above.

Observe that we can also write $J(\mathbf{u}) = J(\hat{\mathcal{U}}_{1:N_t-1})$ for the state transfer. As a matter of fact, *any* unitary control task can be written in this way. What sets either of these tasks apart individually is just how $\hat{\mathcal{U}}_{1:N_t-1}$ appears in the expression — be it gate synthesis, state transfer, observable maximization and whether described by Hilbert space

¹¹ Bangs and similar may have to be smoothed at the edges for better differentiation properties

states or density matrices — and will all thus contain the following principal form for the control derivatives,

$$\text{Gradient} \sim \frac{\partial}{\partial u_n} (\hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1), \quad (4.48a)$$

$$\text{Hessian} \sim \frac{\partial^2}{\partial u_n \partial u_m} (\hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1). \quad (4.48b)$$

The numerical evaluation of these forms is *always* the most expensive part of the numerics. Therefore, even though the considered state transfer formulation leads to “specialized” derivatives they can trivially be extended to cover all other imaginable unitary settings by applying the chain-rule¹².

4.4.1 Propagators and Their Control Landscapes

In addition to discretization, we also critically examine the consequences of choosing a numerical propagator. Why is this important? It is important because any individual choice of propagator, in fact, corresponds to separate control landscapes. We will primarily concern ourselves with the piecewise exact and second-order Suzuki-Trotter (Trotterized) propagators as follows,

$$\hat{\mathcal{U}}_n^{\text{Ex}} = \hat{\mathcal{U}}_n, \quad (4.49a)$$

$$\hat{\mathcal{U}}_n^{\text{ST}_1} = \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} = \hat{\mathcal{U}}_n^{\text{Ex}} + \mathcal{O}(\delta t^3), \quad (4.49b)$$

$$\hat{\mathcal{U}}_n^{\text{ST}_2} = \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} = \hat{\mathcal{U}}_n^{\text{Ex}} + \mathcal{O}(\delta t^3), \quad (4.49c)$$

with the definitions

$$\hat{\mathcal{U}}_n \equiv e^{-i\hat{H}_n \delta t}, \quad \hat{\mathcal{U}}_n^{c/2} \equiv e^{-i\hat{H}_n^c \delta t/2}, \quad \hat{\mathcal{U}}_n^c \equiv \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^{c/2}, \quad \hat{\mathcal{U}}_n^d \equiv e^{-i\hat{H}_n^d \delta t}. \quad (4.49d)$$

where $\hat{H}^d$ and $\hat{H}^c$ are drift and control Hamiltonians as usual. In this section we consider only Schrödinger dynamics, i.e. the interaction-free case $g = 0$. Extension to BEC ($g \neq 0$) is the topic of Ch. 5.

It is pointed out that the operator splitting $\hat{\mathcal{U}}_n^{\text{ST}_1}$ entails the control points both at n and $n + 1$ [216], whereas the $\hat{\mathcal{U}}_n^{\text{ST}_2}$ splitting scheme is fully local in n with the same error [275]. Both versions of the Trotterization accumulates the $\mathcal{O}(\delta t^3)$ error over the full time interval $[0, T]$ and thus yields a $\mathcal{O}(\delta t^2)$ error in total ($N\delta t^3 = T/(\delta t)\delta t^3 = T\delta t^2 \propto \delta t^2$). In other words, we can write

$$J^{\text{ST}}(\mathbf{u}) = J^{\text{Ex}}(\mathbf{u}) + \mathcal{O}_{\text{ST}}(\delta t^2), \quad (4.50)$$

where $J^{\text{Ex}}(\mathbf{u})$ and $J^{\text{ST}_{1,2}}(\mathbf{u})$ are the exact and Trotterized landscapes induced by forming Eq. 4.47 with the piecewise exact propagator $\hat{\mathcal{U}}^{\text{Ex}}$

¹² Alternatively, it is also easy to show that, e.g., a unitary gate is equivalent to a number of simultaneous state transfers [274].

and Trotter propagators $\hat{\mathcal{U}}^{\text{ST}_{1,2}}$, respectively. From this we can clearly see that both the choice of propagator and δt influences the associated landscape. They are all related in the sense that they coincide exactly in the limit $\delta t \rightarrow 0$. In numerical application δt is always finite and therefore we must settle for some $\delta t \leq \tilde{\delta t}$ where the landscape, equivalently dynamical, errors are relatively negligible perturbations.

Although this is an illustrious way of thinking, the consequences of (4.50) are so far trivial. It is indeed of no surprise that choosing an approximant for the time-evolution allows faster evaluation of J at the expense of introducing a controllable error, here $\mathcal{O}_{\text{JST}}(\delta t^2)$, which we purposefully ascribe a label in referral to the dynamics. The real and more subtle consequence is that the derivatives in Eq. (4.48) become remarkably different depending on the chosen propagator. As previously mentioned, it is helpful to recall the same parallel in our discussions of Sec. 3.8 that the efficacy of theoretical open-loop results is tied to how well the underlying model aligns with the experimental reality. Much the same way, in choosing to optimize in any given control landscape — which may or may not represent *exact dynamics* — we must then accordingly always prefer using the “matching” *exact derivatives* lest we similarly induce misalignment and hamper the acquisition and quality of results. In other words, we are now dealing with two distinct error types: (1) that in the dynamics $\mathcal{O}_J$, and (2) that in the derivatives $\mathcal{O}_{\nabla J}$. Importantly, regardless of the former, we should always seek to avoid the latter.

4.4.2 Analytically Exact Derivatives and Problem Representation

We will now derive the analytically exact control derivatives for each of the proposed propagators. To this end, we define an auxiliary state¹³ $|\chi_{N_i}\rangle = |\psi_{\text{tgt}}\rangle$, with $|\chi_n\rangle = \hat{\mathcal{U}}_n^\dagger |\chi_{n+1}\rangle$, and note the following convenient rewritings of the fidelity cost and its derivatives

$$J = \frac{1}{2}(1 - o^*o), \quad (4.51a)$$

$$\frac{\partial J}{\partial u_n} = -\Re \left(o^* \frac{\partial o}{\partial u_n} \right), \quad (4.51b)$$

$$\frac{\partial^2 J}{\partial u_n \partial u_m} = -\Re \left(\left(\frac{\partial o}{\partial u_m} \right)^* \frac{\partial o}{\partial u_n} + o^* \frac{\partial^2 o}{\partial u_n \partial u_m} \right), \quad (4.51c)$$

¹³ It is remarked that this definition is separate from the Lagrange multiplier in the previous section. The same symbol is used since they, indeed, turn out to capture similar information, but in this treatment the motivation of it and other related quantities is simply a notational convenience.

with the overlap o (transfer amplitude) and its derivatives being

$$o = \langle \chi_{N_t} | \psi_{N_t} \rangle = \langle \chi_{N_t} | \hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_n \dots \hat{\mathcal{U}}_1 | \psi_1 \rangle, \quad (4.51d)$$

$$\frac{\partial o}{\partial u_n} = \left\langle \chi_{N_t} \left| \frac{\partial}{\partial u_n} (\hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_n \dots \hat{\mathcal{U}}_1) \right| \psi_1 \right\rangle, \quad (4.51e)$$

$$\frac{\partial^2 o}{\partial u_n \partial u_m} = \left\langle \chi_{N_t} \left| \frac{\partial^2}{\partial u_n \partial u_m} (\hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_n \dots \hat{\mathcal{U}}_1) \right| \psi_1 \right\rangle. \quad (4.51f)$$

The overlap derivatives now contain the forms (4.48) as anticipated.

Their further evaluation depend on the choice of propagator in the following. The gradient calculation steps are shown in some detail while the Hessians are simply stated — I give a full derivation of these in Ref. [3]. For these expressions it is convenient to define the “anti”-Kronecker delta

$$\delta_{i,j} = (1 - \delta_{i,j}). \quad (4.52)$$

We will also make extensive use of the following results from Ref. [3]. Define for notational brevity

$$\hat{H}_n^{c'} \equiv \frac{\partial \hat{H}_n}{\partial u_n} = \frac{\partial \hat{H}_n^c}{\partial u_n}, \quad (4.53)$$

i.e. the prime notation is in referral to the control with time index n and the recursive commutator $[\cdot, \cdot]_k$ as

$$[X, Y]_k = [X, [X, Y]_{k-1}], \quad [X, Y]_0 = Y. \quad (4.54)$$

We can then write the local propagator derivative on close analytical form,

$$\frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} = \hat{\mathcal{U}}_n(-i\delta t) \sum_{k=0}^{\infty} \frac{i^k \delta t^k}{(k+1)!} [\hat{H}_n, \hat{H}_n^{c'}]_k. \quad (4.55)$$

The recursive commutator is not guaranteed to terminate and the summation must therefore be numerically carried out to machine precision ($\epsilon_{\text{mach}} = 2.22 \cdot 10^{-16}$ for the double-precision floating-point format) to be exact in a finite arithmetic. Note that if $\hat{H}_n$ and $\hat{H}_n^{c'}$ already *did* commute then every term $k > 0$ vanishes identically. Nevertheless, in writing out $[\hat{H}_n, \hat{H}_n^{c'}] = [\hat{H}_n^c, \hat{H}_n^{c'}] + [\hat{H}_n^c, \hat{H}_n^{c'}]$ the first term will virtually never commute in practice¹⁴. The second term, however, will often commute. In fact, they are *guaranteed* to do so when represented in a basis that diagonalizes $\hat{H}_n^c$ since two diagonal matrices always commute and therefore $[\hat{H}_n^c, \hat{H}_n^{c'}]_k = \hat{H}_n^{c'} \cdot \delta_{0,k}$. Inspired by this we can write the corresponding derivative for the control

¹⁴ If they do, then the control Hamiltonian is only able to impart extra phase adjustments in their simultaneous basis and not drive transitions. See also the related discussion on the benefits of Hamiltonian non-commutativity in Sec. 3.8.

propagator $\hat{\mathcal{U}}^{c/2}$ represented in the its diagonal basis in anticipation of the following calculations,

$$\frac{\partial \hat{\mathcal{U}}_n^c}{\partial u_n} = \hat{\mathcal{U}}_n^c(-i\delta t) \sum_{k=0}^{\infty} \frac{i^k \delta t^k}{(k+1)!} [\hat{H}_n^c, \hat{H}_n^{c'}]_k = (-i\delta t) \hat{\mathcal{U}}_n^c \hat{H}_n^{c'}, \quad (4.56)$$

and similar for $\hat{\mathcal{U}}^{c/2}$ but with an additional factor $1/2$. Here, $k > 0$ does indeed vanish identically by construction and we remark $[\hat{\mathcal{U}}_n^c, \hat{H}_n^{c'}] = 0$.

Exact Propagator Derivatives

We take $\hat{\mathcal{U}}_n = \hat{\mathcal{U}}_n^{\text{Ex}}$ in Eqs. (4.51). The propagator only depends on the local control u_n and we can therefore write

$$\frac{\partial o^{\text{Ex}}}{\partial u_n} = \left\langle \chi_{N_t} \left| \frac{\partial}{\partial u_n} \left(\hat{\mathcal{U}}_{N_t-1}^{\text{Ex}} \dots \hat{\mathcal{U}}_n^{\text{Ex}} \dots \hat{\mathcal{U}}_1^{\text{Ex}} \right) \right| \psi_1 \right\rangle = \left\langle \chi_{n+1} \left| \frac{\partial \hat{\mathcal{U}}_n^{\text{Ex}}}{\partial u_n} \right| \psi_n \right\rangle. \quad (4.57)$$

Inserting Eq. (4.55) and returning to Eqs. (4.51) we obtain the exact gradient

$$\frac{\partial J^{\text{Ex}}}{\partial u_n} = \Re \left(i o^* \left\langle \chi_n \left| \left(\sum_{k=0}^{\infty} \frac{i^k \delta t^k}{(k+1)!} [\hat{H}_n, \hat{H}_n^{c'}]_k \right) \right| \psi_n \right\rangle \right) \delta t \quad (4.58)$$

$$= \Re \left(i o^* \left\langle \chi_n \left| \hat{H}_n^{c'} \right| \psi_n \right\rangle \right) \delta t + \mathcal{O}_{\nabla J^{\text{Ex}}}(\delta t^2). \quad (4.59)$$

for $1 \leq n \leq N_t$ with a further convenient definition for the “infinite tail”

$$\mathcal{O}_{\nabla J^{\text{Ex}}}(\delta t^2) = \Re \left(i o^* \left\langle \chi_n \left| \left(\sum_{k=1}^{\infty} \frac{i^k \delta t^k}{(k+1)!} [\hat{H}_n, \hat{H}_n^{c'}]_k \right) \right| \psi_n \right\rangle \right) \delta t. \quad (4.60)$$

In a similar way we can calculate the exact Hessian elements $n \geq m$ (without loss of generality due to $n \rightleftharpoons m$ symmetry) for the exact propagator

$$\begin{aligned} \frac{\partial^2 J^{\text{Ex}}}{\partial u_n \partial u_m} &= -\Re \left(\langle \psi_m | \frac{\partial \hat{\mathcal{U}}_m}{\partial u_m} | \chi_{m+1} \rangle \langle \chi_{n+1} | \frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} | \psi_n \rangle \right) \\ &\quad - \Re \left(o^* \langle \chi_{n+1} | \frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} \left(\prod_{j=m+1}^{n-1} \hat{\mathcal{U}}_j \right) \frac{\partial \hat{\mathcal{U}}_m}{\partial u_m} | \psi_m \rangle \right) \delta_{n,m} \\ &\quad - \Re \left(o^* \langle \chi_{n+1} | \frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2} | \psi_n \rangle \right) \delta_{n,m}. \end{aligned} \quad (4.61)$$

We yet have to calculate $\frac{\partial^2 \hat{\mathcal{U}}_n}{\partial u_n^2}$ and, although straightforward, the result is somewhat cumbersome because it requires derivatives of the recursive commutator $[\hat{H}_n, \hat{H}_n^{c'}]_k'$, see Ref. [3].

BOUNDARY BEHAVIOR — These expressions are valid for all $n, m = 1 : N_t - 1$. Elements involving the last time index N_t are zero because there is no $\hat{\mathcal{U}}_{N_t}$ propagator. This endpoint behavior could be encompassed explicitly by multiplying gradient elements with $E_n^{\text{Ex}} = \delta_{n, N_t}$ and Hessian elements with $E_n^{\text{Ex}} \cdot E_m^{\text{Ex}}$.

Trotterized Propagator Gradients

We now take $\hat{\mathcal{U}}_n = \hat{\mathcal{U}}_n^{\text{ST}_{1,2}}$ in Eqs. (4.51). The overlap derivatives are contracted similarly to the exact case

$$\frac{\partial o^{\text{ST}_1}}{\partial u_n} = \left\langle \chi_{n+1} \left| \frac{\partial}{\partial u_n} \left(\hat{\mathcal{U}}_n^{\text{ST}_1} \hat{\mathcal{U}}_{n-1}^{\text{ST}_1} \right) \right| \psi_{n-1} \right\rangle, \quad (4.62a)$$

$$\frac{\partial o^{\text{ST}_2}}{\partial u_n} = \left\langle \chi_{n+1} \left| \frac{\partial \hat{\mathcal{U}}_n^{\text{ST}_2}}{\partial u_n} \right| \psi_n \right\rangle, \quad (4.62b)$$

respectively due to their locality in either n and $n + 1$ or just n .

For $\text{ST}_{1,2}$ and using the diagonality and commutativity criteria (4.56) we find

$$\begin{aligned} \frac{\partial}{\partial u_n} \left(\hat{\mathcal{U}}_n^{\text{ST}_1} \hat{\mathcal{U}}_{n-1}^{\text{ST}_1} \right) &= \frac{\partial}{\partial u_n} \left(\left(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \right) \left(\hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \right) \right) \\ &= \left(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \right) \frac{\partial \hat{\mathcal{U}}_n^c}{\partial u_n} \left(\hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \right) \\ &= \left(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \right) \left\{ \hat{\mathcal{U}}_n^c (-i\delta t \hat{H}_n^{c'}) \right\} \left(\hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \right) \\ &= (-i\delta t) \cdot \left(\hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \right) \hat{H}_n^{c'} \left(\hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_{n-1}^d \hat{\mathcal{U}}_{n-1}^{c/2} \right) \\ &= (-i\delta t) \cdot \hat{\mathcal{U}}_n^{\text{ST}_1} \hat{H}_n^{c'} \hat{\mathcal{U}}_{n-1}^{\text{ST}_1}, \end{aligned} \quad (4.63a)$$

$$\begin{aligned} \frac{\partial \hat{\mathcal{U}}_n^{\text{ST}_2}}{\partial u_n} &= \frac{\partial \hat{\mathcal{U}}_n^{c/2}}{\partial u_n} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} + \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \frac{\partial \hat{\mathcal{U}}_n^{c/2}}{\partial u_n} \\ &= -\frac{i\delta t}{2} \left(\hat{\mathcal{U}}_n^{c/2} \hat{H}_n^{c'} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} + \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \hat{H}_n^{c'} \right) \\ &= -\frac{i\delta t}{2} \left(\hat{H}_n^{c'} \hat{\mathcal{U}}_n^{\text{ST}_2} + \hat{\mathcal{U}}_n^{\text{ST}_2} \hat{H}_n^{c'} \right). \end{aligned} \quad (4.63b)$$

Substituting these expressions into the overlap derivatives and returning to Eqs. (4.51) we obtain the exact gradients

$$\frac{\partial f^{\text{ST}_1}}{\partial u_n} = \Re \left(i o^* \left\langle \chi_n \left| H_n^{c'} \right| \psi_n \right\rangle \right) \delta t, \quad (4.64a)$$

$$\frac{\partial f^{\text{ST}_2}}{\partial u_n} = \frac{1}{2} \sum_{p=n}^{n+1} \Re \left(i o^* \left\langle \chi_p \left| \hat{H}_n^{c'} \right| \psi_p \right\rangle \right) \delta t \quad (4.64b)$$

The exact Hessian elements $n \geq m$ (without loss of generality due to $n \rightleftharpoons m$ symmetry) for the Trotter propagators are

$$\begin{aligned} \frac{\partial^2 J^{\text{ST}_1}}{\partial u_n \partial u_m} = & -\Re \left(\langle \psi_m | \hat{H}_m^{c'} | \chi_m \rangle \cdot \langle \chi_n | \hat{H}_n^{c'} | \psi_n \rangle \right) \delta t^2 \\ & + \Re \left(o^* \langle \chi_n | \hat{H}_n^{c'} \left(\prod_{j=m+1}^n \hat{\mathcal{U}}_j \right) \hat{H}_m^{c'} | \psi_m \rangle \right) \delta t^2 \cdot \delta_{n,m} \\ & + \Re \left(io^* \langle \chi_n | \left(\hat{H}_n^{c''} - i\delta t (\hat{H}_n^{c'})^2 \right) | \psi_n \rangle \right) \delta t \cdot \delta_{n,m}, \quad (4.65a) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 J^{\text{ST}_2}}{\partial u_n \partial u_m} = & + \Re \left(\frac{1}{4} \sum_{p=n}^{n+1} \sum_{q=m}^{m+1} \langle \psi_q | \hat{H}_m^{c'} | \chi_q \rangle \langle \chi_p | \hat{H}_n^{c'} | \psi_p \rangle \right) \delta t^2 \\ & + \Re \left(\frac{o^*}{4} \sum_{p=n}^{n+1} \sum_{q=m}^{m+1} \left\langle \chi_p \left| \hat{H}_n^{c'} \prod_{j=q}^{p-1} \hat{\mathcal{U}}_n^{\text{ST}_2} \hat{H}_m^{c'} \right| \psi_q \right\rangle \right) \delta t^2 \delta_{n,m} \\ & + \Re \left(\frac{io^*}{2} \left\{ \sum_{j=n}^{n+1} \left\langle \chi_j \left| \left(\hat{H}_n^{c''} - \frac{i\delta t}{2} (\hat{H}_n^{c'})^2 \right) \right| \psi_j \right\rangle \right. \right. \\ & \quad \left. \left. - i\delta t \left\langle \chi_{n+1} \left| \hat{H}_n^{c'} \hat{\mathcal{U}}_n^{\text{ST}_2} \hat{H}_n^{c'} \right| \psi_n \right\rangle \right\} \right) \delta t \delta_{n,m}. \quad (4.65b) \end{aligned}$$

and no further reductions remain.

BOUNDARY BEHAVIOR — The ST_1 expressions are valid for bulk indices $n, m = 2 : N_t - 1$ due to Eqs. (4.63a)–(4.63b). At the end points 1 and N_t there is only one control propagator in Eq. (4.63a) and thus only, comparatively, “half” the derivative contribution. This could explicitly be accounted for in the expressions by multiplying the gradient elements with $E_n^{\text{ST}_1} = 1 - \frac{1}{2}(\delta_{1,n} + \delta_{N_t,n})$ and the Hessian elements with $E_n^{\text{ST}_1} \cdot E_m^{\text{ST}_1}$.

The ST_2 expressions are valid for indices $n = 1 : N_t - 1$. The elements for final index N_t are zero because there is no $\hat{\mathcal{U}}_{N_t}$ propagator. This could explicitly be accounted for in the expressions by multiplying the gradient elements with $E_n^{\text{ST}_2} = \delta_{N_t,n}$ and the Hessian elements with $E_n^{\text{ST}_2} \cdot E_m^{\text{ST}_2}$.

4.4.3 Control Parametrization

It is trivial to extend the results in this setting to any control parametrization $u_n = u(t_n; \theta)$ where $\theta \in \mathbb{R}^M$ are the control parameters as in Sec. 3.5. Since J is a function of all $\{u_n(\theta)\}_{n=1}^{N_t}$ which are each possibly

functions of each control parameters $(\theta)_i = \theta_i$ we readily apply the multivariate chain rule and a few lines of evaluation

$$\frac{\partial J}{\partial \theta_i} = \sum_{n=1}^{N_t} \frac{\partial J}{\partial u_n} \cdot \frac{\partial u_n}{\partial \theta_i}, \quad (4.66a)$$

$$\frac{\partial^2 J}{\partial \theta_i \partial \theta_j} = \sum_{n=1}^{N_t} \frac{\partial J}{\partial u_n} \cdot \frac{\partial^2 u_n}{\partial \theta_i \partial \theta_j} + \sum_{n,m=1}^{N_t} \frac{\partial^2 J}{\partial u_n \partial u_m} \cdot \frac{\partial u_n}{\partial \theta_i} \cdot \frac{\partial u_m}{\partial \theta_j}, \quad (4.66b)$$

for any landscape J (equivalently propagator). Derivatives of u_n are trivial since they are on closed analytical form and the main computational effort remains fully in the derivatives of J . This structure coincides with that found in Sec. 4.3.6 and the examples from there carries trivially over to here.

4.4.4 Constraint Costs

In handling constraint costs such as those of Sec. 3.6 in the discretized setting we need to apply appropriate numerical integration and possibly finite differencing stencils. For example, in choosing finite differencing schemes accurate to order $\mathcal{O}(\delta t^2)$ and the simplest uniform sampling integral approximation we have for e.g. the amplitude, soft-bound, and derivative regularization

$$J_\alpha = \frac{\alpha}{2} \delta t \sum_{i=1}^{N_t} u_i^2, \quad (4.67a)$$

$$J_\sigma = \frac{\sigma}{2} \delta t \sum_{i=1}^{N_t} \begin{cases} (u_i - u_{\min})^2, & u_i < u_{\min} \\ 0, & u_{\min} \leq u_i \leq u_{\max} \\ (u_i - u_{\max})^2, & u_i > u_{\max} \end{cases} \quad (4.67b)$$

$$J_\gamma = \frac{\gamma}{8\delta t} \left([-3u_1 + 4u_2 - u_3]^2 + \sum_{i=2}^{N_t-1} (u_{i+1} - u_{i-1})^2 + [3u_{N_t} - 4u_{N_t-1} + u_{N_t-2}]^2 \right), \quad (4.67c)$$

Here, we use a finite differencing with forward and backward approximations for points $n = 1, N_t$, respectively, while the bulk indices $n = 2 : N_t - 1$ uses centered approximations.

In the spirit of discretize-then-optimize it is then these expressions we should explicitly calculate the derivatives of. The amplitude and soft-bound derivatives are easy,

$$\frac{\partial J_\alpha}{\partial u_n} = \alpha \delta t u_n, \quad \frac{\partial^2 J_\alpha}{\partial u_m \partial u_n} = \alpha \delta t \delta_{n,m}, \quad (4.68)$$

$$\frac{J_\sigma}{\partial u_n} = \sigma \delta t \begin{cases} (u_n - u_{\min}) \\ 0 \\ (u_n - u_{\max}) \end{cases}, \quad \frac{\partial^2 J_\sigma}{\partial u_m \partial u_n} = \sigma \delta t \delta_{n,m} \begin{cases} 1 \\ 0 \\ 1 \end{cases}. \quad (4.69)$$

For the derivative regularization, the calculations are a bit more tedious — of course this is self-inflicted lamentation as we could also have chosen simpler $\mathcal{O}(\delta t)$ stencil approximations. The resulting gradient is

$$\nabla J_\gamma = \frac{\gamma}{4\delta t} \begin{bmatrix} 10u_1 - 12u_2 + 2u_3 \\ -12u_1 + 17u_2 - 4u_3 - u_4 \\ 2u_1 - 4u_2 + 3u_3 - u_5 \\ \vdots \\ 2u_n - u_{n-2} + u_{n+2} \\ \vdots \\ 2u_{N_t} - 4u_{N_t-1} + 3u_{N_t-2} - u_{N_t-4} \\ -12u_{N_t} + 17u_{N_t-1} - 4u_{N_t-2} - u_{N_t-3} \\ 10u_{N_t} - 12u_{N_t-1} + 2u_{N_t-2} \end{bmatrix}, \quad (4.70)$$

and “ $\vdots$ ” signify extension over the bulk points (i.e. the first and last three elements are special). This structure carries over to the Hessian which can be written in vector-stacking notation (subscripts denote rows and horizontal bars are simply visual separators)

$$\nabla^2 J_\gamma = \begin{bmatrix} [\nabla^2 J_\gamma]_{1:3} \\ [\nabla^2 J_\gamma]_{4:N_t-3} \\ [\nabla^2 J_\gamma]_{N_t-2:N_t'} \end{bmatrix} \quad (4.71)$$

with the matrices evaluating to

$$[\nabla^2 J_\gamma]_{1:3} = \frac{\gamma}{4\delta t} \begin{bmatrix} 10 & -12 & 2 & 0 & 0 & 0 & \dots \\ -12 & 17 & -4 & -1 & 0 & 0 & \dots \\ 2 & -4 & 3 & 0 & -1 & 0 & \dots \end{bmatrix}, \quad (4.72a)$$

$$[\nabla^2 J_\gamma]_{4:N_t-3} = \frac{\gamma}{4\delta t} \begin{bmatrix} 0 & -1 & 0 & 2 & 0 & -1 & 0 & 0 & \dots \\ & & \ddots & & \ddots & & \ddots & & \\ \dots & 0 & 0 & -1 & 0 & 2 & 0 & -1 & 0 \end{bmatrix}, \quad (4.72b)$$

$$[\nabla^2 J_\gamma]_{N_t-2:N_t'} = \frac{\gamma}{4\delta t} \begin{bmatrix} \dots & 0 & -1 & 0 & 3 & -4 & 2 \\ \dots & 0 & 0 & -1 & -4 & 17 & -12 \\ \dots & 0 & 0 & 0 & 2 & -12 & 10 \end{bmatrix}, \quad (4.72c)$$

and the dots denote continuation of the number they point to.

These are the exact derivatives for amplitude, soft-bound, and derivative regularization for the $\mathcal{O}(\delta t^2)$ finite differencing stencil with simple uniform integration sampling.

4.5 COMPARISON OF DERIVATIVES

We have now calculated several derivatives through optimize-then-discretize and discretize-then-optimize approaches. For the latter we have also considered several distinct propagators and we proceed to compactly summarize these and their associated gradients for all $n = 1 : N_t$ (for the $g = 0$ case) by writing

$$\hat{\mathcal{U}}_n = \mathcal{T} e^{-i \int_{t_n}^{t_n + \delta t} \hat{H}(t) dt} \rightarrow \frac{\partial \bar{J}}{\partial u_n} = \delta t \cdot \Omega_n^n, \quad (4.73a)$$

$$\hat{\mathcal{U}}_n^{\text{Ex}} = e^{-i \hat{H}_n \delta t} \rightarrow \frac{\partial J^{\text{Ex}}}{\partial u_n} = \delta t \cdot \Omega_n^n + \mathcal{O}_{\nabla J^{\text{Ex}}}(\delta t^2), \quad (4.73b)$$

$$\hat{\mathcal{U}}_n^{\text{ST}_1} = \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2} \rightarrow \frac{\partial J^{\text{ST}_1}}{\partial u_n} = \delta t \cdot \Omega_n^n, \quad (4.73c)$$

$$\hat{\mathcal{U}}_n^{\text{ST}_2} = \hat{\mathcal{U}}_n^{c/2} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2} \rightarrow \frac{\partial J^{\text{ST}_2}}{\partial u_n} = \delta t \cdot \frac{(\Omega_n^n + \Omega_{n+1}^{n+1})}{2}. \quad (4.73d)$$

The control and drift propagators of $\hat{\mathcal{U}}_n^{\text{ST}}$ are given in Eq. (4.49d) as usual and it is important to emphasize that the would-be infinite $\mathcal{O}_{\nabla J^{\text{ST}}}$ tail is eliminated *only* when representing the problem in the control-diagonal basis. Additionally, we have used the following convenient definitions for first-order elements Ω_n^p , the infinite tail $\mathcal{O}_{\nabla J^{\text{Ex}}}$, and each gradient element is implicitly multiplied by the appropriate end point modifications E_n ,

$$\Omega_n^p = \Re \left(i o^* \left\langle \chi_p \left| H_n^{c'} \right| \psi_p \right\rangle \right), \quad (4.74a)$$

$$\mathcal{O}_{\nabla J^{\text{Ex}}}(\delta t^2) \propto \left\langle \chi_n \left| \left(\sum_{k=1}^{\infty} \frac{i^k \delta t^k}{(k+1)!} [\hat{H}_n, \hat{H}_n^{c'}]_k \right) \right| \psi_n \right\rangle \delta t, \quad (4.74b)$$

$$\bar{E}_n = \delta_{n,1} + \delta_{n,N_t}, \quad (4.74c)$$

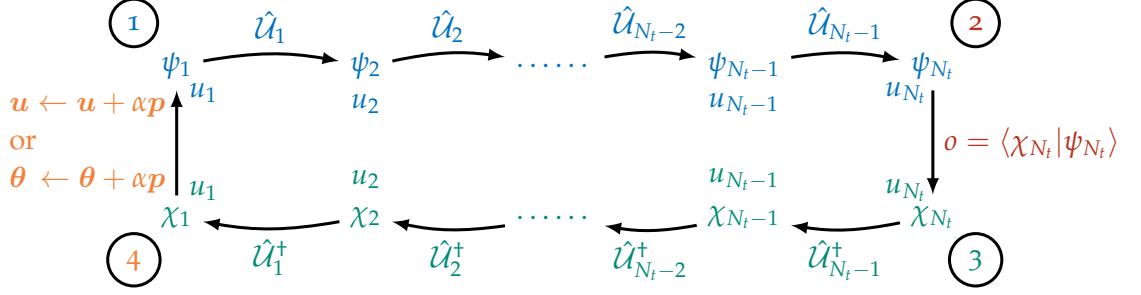
$$E_n^{\text{ST}_1} = 1 - \frac{1}{2}(\delta_{1,n} + \delta_{N_t,n}), \quad (4.74d)$$

$$E_n^{\text{Ex}} = E_n^{\text{ST}_2} = \delta_{n,N_t}. \quad (4.74e)$$

With these equations, coupled with the content described in Fig. 4.1 and possibly converting to a particular control parametrization Eq. (4.66), we can now perform iterative derivative-based optimization with the common algorithmic structure shown in Fig. 4.4.

There are several noteworthy discussion and comparative points associated with Eqs. (4.73). These are treated in turn through the following subsections. As a primer, note that all propagators and derivatives correctly coincide in the limit $\delta t \rightarrow 0$ as we would expect (differences in end point conditions also become negligible). For finite δt , however, we see that the derivatives have a fundamentally different form and, more critically, also scale differently with δt . We focus here on the gradients for brevity but similar yet ultimately more exaggerated differences are found for the Hessians.

Figure 4.4: Adapted from the qualifying exam progress report. Optimization loop (iteration number k suppressed).



- ①–② Forward propagate $|\psi_{\text{ini}}\rangle = |\psi_1\rangle \rightarrow |\psi_{N_t}\rangle$ with $\mathbf{u}^{(k)}$.
- ②–③ Calculate $o = \langle \chi_{N_t} | \psi_{N_t} \rangle = \langle \chi_{N_t} | \hat{U}_{1:N_t-1} | \psi_1 \rangle$.
- ③–④ Backward propagate $|\psi_{\text{tgt}}\rangle = |\chi_{N_t}\rangle \rightarrow |\chi_1\rangle$ with $\mathbf{u}^{(k)}$.
- ④–① Calculate the gradient $\nabla J^{(k)}$ (and possibly Hessian $\nabla^2 J^{(k)}$), search direction $\mathbf{p}^{(k)}$, and step size $\alpha^{(k)}$. Take the step $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha^{(k)} \mathbf{p}^{(k)}$ or $\theta^{(k+1)} = \theta^{(k)} + \alpha^{(k)} \mathbf{p}^{(k)}$. Check termination conditions — continue to iteration $k + 1$?

It is cautioned that the similarity between ∇J^{ST_1} and $\overline{\nabla J}$ is more subtle than it seems and this is also discussed below.

4.5.1 The Role of Exact Derivatives in Action

Let us consider first the discretize-then-optimize results. Here, the control-diagonal Trotter gradients are already *exact* with just the first-order δt term whereas the piecewise exact propagator requires additional evaluation of an expensive infinite recursive commutator sum (truncated to some $k \leq k_{\text{max}}$ in finite arithmetic) to achieve a similarly exact gradient. Recall that the initial motivation for introducing $\hat{U}_n^{\text{ST}}$ was to accelerate the numerical procedure for solving the quantum dynamical equations of motion since $\hat{U}_n^{\text{Ex}}$ numerically corresponds to an expensive matrix exponentiation. Among other attractive properties, the Trotter-splitting is amenable to computational speed-up and caching techniques [3] when described in the control-diagonal basis. Now, more profoundly, it has further turned out — contingent on the *same* control-diagonal representation — that this particular type of approximant is also significantly less encumbered in attaining exact derivatives: the runtime evaluation thereof scales only with the time it takes to solve the dynamics itself ($2N_t$ propagations) which by virtue of the preceding sentence is already very fast. We therefore enjoy significant numerical complexity reductions both with respect

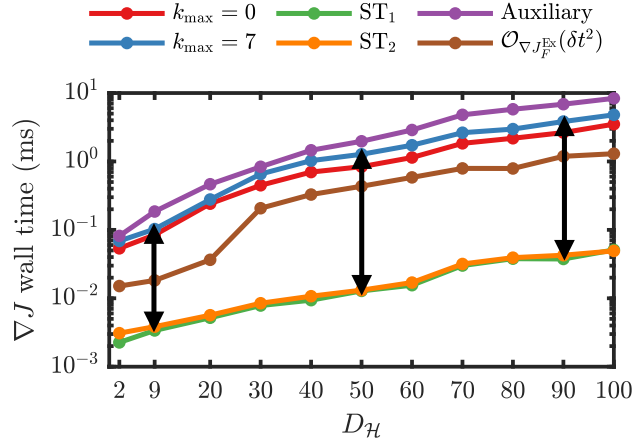


Figure 4.5: Adapted from Ref. [3]. Gradient and commutator tail (Eqs. (4.73)–(4.74)) calculation time as a function of Hilbert space dimensionality $D_{\mathcal{H}}$. Here, $k_{\max} = 0$ and $k_{\max} = 7$ are the first-order approximation and exact gradients for J^{Ex} , respectively, and their wall time difference is due to the commutator tail (summed to machine precision). In ensuring exactness, the $^{\text{ST}}$ gradients have beneficial scaling properties as evidenced by the exponentially increasing performance gap roughly indicated by the arrows. The gap is aggregated by individual contributions from the dynamics (trivially) and the tail computation (non-trivially). “Auxiliary” refers to an alternative exact derivative calculation through the *auxiliary matrix method* [263, 276, 277].

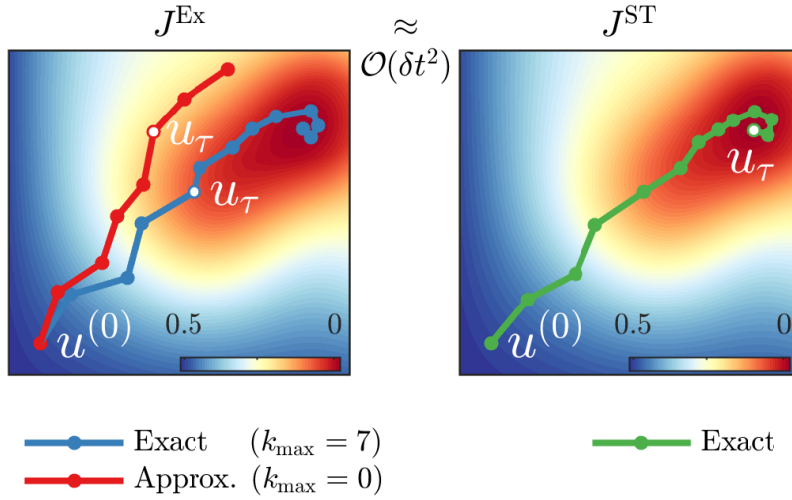


Figure 4.6: Adapted from Ref. [3]. Ascending the fidelity control landscape with local optimization either in the piecewise exact J^{Ex} landscape (left) or in one of the Trotter landscape J^{ST} (right). The dotted lines denote optimization trajectories starting from the same seed $u^{(0)}$ when utilizing different of derivatives. Approximate derivatives lead one (usually irrecoverably) astray. Exact derivatives yields guaranteed monotonic decrease in cost per iteration. However, the per-iteration wall time performance gap of Fig. 4.5 manifests as a performance gap in the achievable results (white dots) for the same, fixed optimization wall time τ .

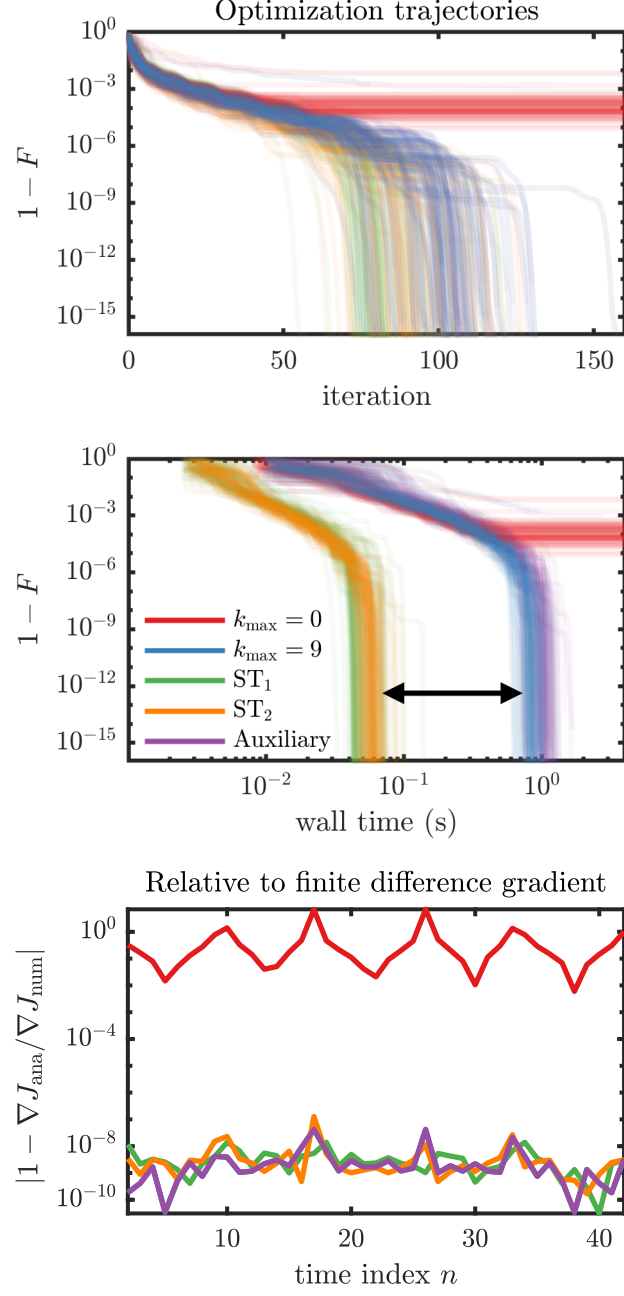


Figure 4.7: Adapted from Ref. [3]. Demonstration of optimizing in exact and Trotterized landscapes for the two-level LZ problem from Fig. 3.3. Top and middle: We optimize with the different gradients using p_{BFGS} for the same 100 (piecewise constant) uniform randomly generated seeds, $u_n = \text{uniform}(-10, 10)$, at $T = 1.01\pi \gtrsim T_{\min}$ with $\delta t = 0.075$. In this example $k_{\max} = 9$ is necessary to sum to machine precision. The performance gap between the exact methods is roughly denoted by the horizontal arrow. Bottom: Analytical gradients (∇J_{ana}) relative to their numerical central finite difference approximation (∇J_{num}) with perturbation $\epsilon_{\text{pert}} = \epsilon_{\text{mach}}^{1/3}$ for a constant example control with $u_n = 5$. Note that the Auxiliary and $k_{\max} = 9$ gradients are identical to machine precision and thus yield (nearly) the same iteration trajectories, and, incidentally, nearly the same wall time trajectories.

to dynamics and ensuring gradient exactness in comparison to the piecewise exact propagator.

The resulting performance hierarchy of the herein discussed propagators are numerically verified as a function of Hilbert space size $D_{\mathcal{H}}$ in Fig. 4.5 and the conceptual consequences thereof shown in Fig. 4.6. As a baseline example for the “simplest non-simple” case $D_{\mathcal{H}} = 2$, Fig. 4.7 shows how this performance gap manifests numerically in a large scale optimizational setting. Concretely, utilizing approximate derivatives severely limits the achievable fidelity (the BFGS approximation is scrambled). Utilizing on the other hand any of the exact derivatives and associated propagators eventually lead to machine-precision optimal results in this example. However, while requiring roughly the same number of iterations, there is more than an order of magnitude difference in the required wall time to achieve this. In other words, it is always much faster to maneuver the control-diagonal Trotter landscape and the benefit increases monotonically with Hilbert space size $D_{\mathcal{H}}$ (see e.g. Ref. [3] for another concrete example with $D_{\mathcal{H}} = 9$ where a rotation into the control diagonal basis is also required). Extrapolating $D_{\mathcal{H}}$ to the extreme, this aggregate computational performance provides a scaffolding for realizing efficient exact derivative-based optimization in very high-dimensional many-body systems in the high-fidelity limit. This is the topic of Ch. 7 and Ref. [4] where we consider a system far beyond the capabilities of exact diagonalization approaches, necessitating a matrix product state description.

The Size of δt

The exact gradient for the exact piecewise constant propagator $\hat{\mathcal{U}}^{\text{Ex}}$ depends on the infinite tail Eq. 4.74b. The number of terms needed for machine precision in the infinite tail depends on the size of δt — the smaller δt , the fewer terms k_{max} necessary. However, simply reducing δt also increases the number of time steps $N_t = \lfloor T/\delta t \rfloor + 1$ accordingly and this means by Fig. 4.5 that the now more efficiently calculable tail comes at the expense of significantly increased dynamical integration efforts. Thus, reducing δt does not mend the aggregate performance discrepancy between $\hat{\mathcal{U}}^{\text{Ex}}$ and $\hat{\mathcal{U}}^{\text{ST}}$ and the latter are still preferable.

More to this point, the $\hat{\mathcal{U}}^{\text{ST}}$ gradients are exact to first order in δt no matter its actual size. Among others, this has intrinsic benefits for some of the advanced concepts described in Sec. 3.7, namely ensemble robustness and homotopy methodologies.

Ensemble optimization requires per iteration simultaneous evaluation of a number of L gradients, typically tens or so, and any analytical inexactness is thus compounded more severely. Implementing the necessary exactness scales the gap in e.g. Fig. 4.5 by L making $\hat{\mathcal{U}}^{\text{ST}}$ even more attractive.

In homotopy methodologies, iteration speed ups are achieved by initially choosing a set of problem parameters that are not necessarily dynamical exact. With the discussions in this chapter we now interpret this as optimizing in distinct control landscapes where we prefer corresponding exact derivatives. A standard homotopy parameter is δt , initially chosen to be a relatively large and then sequentially reduced to a physically reasonable one. As before, however, derivative exactness in the large δt limit makes a huge difference for $\hat{U}^{\text{Ex}}$ by the number of terms k_{max} needed while it is fully inconsequential for $\hat{U}^{\text{ST}}$. This makes δt -homotopies with $\hat{U}^{\text{ST}}$ much more wieldy.

4.5.2 Broadening the Scope

In Ref. [3] I also discuss other approaches and standard propagation schemes (Krylov-Lanczos, finite Taylor expansions, Crank-Nicolson, Chebyshev [7, 278–280]). Including $\hat{U}^{\text{Ex}}$, it turns out that they all suffer from severely scaling computational bottlenecks in evaluating their exact derivatives due to e.g. expensive recursive commutators [271], matrix exponentials [263], matrix diagonalizations [7], and others. It is also remarked in passing that unitary propagator approaches are, in general, preferred over explicit coupled ODE solvers like the Euler and fourth order Runge-Kutta integrators. Despite the latter's simplicity they have several flaws. First, they have poor accuracy and are unstable even for constant $\hat{H}$ unless δt is taken very small [281]. Second, the typical trick in these methods of picking adaptive step sizes does not blend well with our requirements here since e.g. gradients are calculated using $\sim \langle \chi(t_n) | \dots | \psi(t_n) \rangle$ which requires the same time grid for both states¹⁵. Third, they do not preserve the central property of unitarity [282] — some propagator approaches also suffer from this but the ones presented here in detail, $\hat{U}^{\text{Ex}}$ and $\hat{U}^{\text{ST}}$, do preserve unitarity.

Generalizability

A consideration towards the extended success of the control-diagonal Trotter methodology lies in its generalization to multiple controls. I give a full analysis in Ref. [3] and simply summarize the results here. The main potential difficulty does not lie with the multiplicity of controls, but rather the commutation relations of the control Hamiltonians and thus the feasibility of maintaining the control-diagonal criterion. This is because sets of non-commuting control Hamiltonians cannot be simultaneously diagonalized and we must therefore transform between the different diagonal representations to preserve the formulated advantages. Although imposing somewhat of an additional

¹⁵ We can envision adaptive step sizes being boonful for other reasons. They just have to be consistently chosen for both.

computational overhead, it turns out that these transformations are (nearly) always feasible compared to the alternatives. Additionally, the unfeasible edge cases can be eliminated categorically for bilinear systems (which are especially pervasive in the literature), or by sequentially optimizing the individual, mutually commuting sets while, momentarily, considering the rest a drift contribution. In these ways, even the (immediate) limitations encountered for non-commuting control Hamiltonians can be alleviated to a wide extent. Thus, the control-diagonal Trotter methodology generalizes well to multiple controls.

4.5.3 First-Order Coincidences

It is notable that the first-order approximation of ∇J^{Ex} (i.e. $k_{\text{max}} = 0$) coincides with the form of ∇J^{ST_1} but not with ∇J^{ST_2} which is the most popularly used Trotter-form. This property is in fact enjoyed exclusively by ∇J^{ST_1} among all the standard dynamical schemes.

A related interesting result is the seemingly perfect coincidence $\nabla J^{\text{ST}_1} = \overline{\nabla J}$ as obtained separately by discretize-then-optimize and optimize-then-discretize approaches. This is, however, somewhat of an illusion because the latter was derived from the “true” time-ordered exponential implying the continuous limit $\delta t \rightarrow 0$. Upon discretization, time slices $t_n \rightarrow t_n + \delta t$ of the time-ordered exponential naturally becomes the piecewise exact exponential — and comparing $\overline{\nabla J}$ and ∇J^{Ex} we find that the crucial information about the infinite tail has been lost in the process! Nevertheless, if one had indeed derived $\overline{\nabla J}$ and, for one reason or another, felt compelled to precisely use $\hat{U}_n^{\text{ST}_1}$ and a control-diagonal representation, then one would indeed achieve $J^{\text{ST}_1} = \overline{\nabla J}$ in non-illusory terms. The only discrepancy then occurs at the end points, $E_n^{\text{ST}_1} \neq \overline{E}_n$, but this is typically not a problem because $\overline{E}_n$ -type zeroing is often explicitly imposed *a posteriori* due to separate problem considerations regardless of propagator choice.

4.5.4 The Interacting Case, $g \neq 0$

Let us conclude here by recalling that the comparisons in this section have concerned the $g = 0$ case. This is because we have, so far, derived expressions for $g \neq 0$ (the optimality equations (4.38)) only in the optimize-then-discretize setting. We have nevertheless spent a considerable amount of energy convincing ourselves that the results from this approach are less preferable in the limiting $g = 0$ case. The question is, then, how well do Eqs. (4.38) coincide with the corresponding $g \neq 0$ exact control derivatives from the discretize-then-optimize setting? After all, it turned out in a preceding paragraph that certain fortuitous happenstances could lead to the coveted exact derivatives

by “accident”. Perhaps the same mechanisms could extricate us from having to bother with the discretize-then-optimize mindset even for $g \neq 0$? Unfortunately, this turns out to not be the case as shown in the following Ch. 5 where we generalize the calculations.

4.6 PERSPECTIVES

We conclude this chapter by scoping out and reflecting over a larger variety of algorithmic approaches and briefly touch on a number of related emergent acronyms. The prevalent perspective of quantum optimal control, in this thesis, is that it consists of equal parts numerical optimization and simulation of quantum systems. As remarked in Sec. 4.1 we inherit a lot of concepts directly from the formers’ extensive and more mature literature. Specifically, an appropriate “language” and list of methodologies have carried over quite seamlessly. Nevertheless, the organic evolution of the “quantum-part” has lead to a somewhat, at times, confusing terminology. This is clearly somewhat subjective and is not meant with any hint of inimicality. Note that this section is not primarily geared toward experts of the field.

It is easiest to first solidify the stance taken here. We do this by recounting explicit as well as implicit ideas, assumptions, and results from this and the previous chapter. This not only serves as a convenient grounds for comparison, but also a recapitulation of parts of the presented material in a few new words and slightly different light.

4.6.1 What is Quantum in Optimal Control?

The cost function/control landscape J is fundamental. It is uniquely defined in terms of all its problem parameters and this is related to the discussion of homotopy parameters in Sec. 3.7. This includes, e.g., the physical problem to be solved, the available control parameters, and constraints thereof since they define the mathematical function(al) *structure* of J but also includes all the choices of numerical implementation for discretization, time-evolution, and so on, because they define the *details* of J .

Any change, substitution, or perturbation, small or large, to the problem parameters defines another functional J' that may or may not be closely related to J . Here are a few, some perhaps bordering on obvious, examples.

- In changing only the underlying physical problem to be solved, the similarity of J to J' depends on, e.g., whether one is just tweaking the target state slightly or if one is describing a completely different type of unitary task, trapping geometry, or physical system.
- As example from the previous chapter, in changing only the parametrization of the control the similarity of J to J' can vary prodigiously or

not, even if they both are still conceptually representing the same physical problem.

- As example from this chapter, in changing only the time-evolution operator the similarity of J to J' depends on δt even if they both, again, are still conceptually representing the same physical problem.

It is remarked that the control is always represented in a finite-size parametrization in the numerical setting presented here, but the values themselves are essentially continuous (down to the machine precision). The maximum number of independent parameters is equal to the number of time steps δt (full resolution piecewise constant parametrization).

Having cemented the distinctness of every possible J , we can now turn to locating its troughs which we associate with optimal controls. We can do this in an innumerable number of ways, but drawing on the wisdom of extensive nonlinear optimization literature [214, 245, 260–262] we are lead to methodologies that fundamentally rely on locally evaluating the cost function and possibly one or both of its first two derivatives — these “Geometric” (“true”) quantities, $G = \{J^G, \nabla J^G, \nabla^2 J^G\}$, are again uniquely fixed by considering all the computational problem parameters. That is, G exists “out there” whether we are able to calculate it by whatever means or not. For the best optimization results, it is then incumbent on us to try and procure Computed quantities, $C = \{J^C, \nabla J^C, \nabla^2 J^C\}$, that to the best possible approximation is aligned with the geometric ones, $C \approx G$, and preferably are exactly equal to machine precision, $C = G$:

$$\text{(computed)} \quad \{J^C, \nabla J^C, \nabla^2 J^C\} = \{J^G, \nabla J^G, \nabla^2 J^G\} \quad \text{(geometric)} \quad (4.75)$$

The choice of optimizer determines how many elements of the tuple C need be calculated.

It is always trivial to exactly compute the geometric cost J^G because it is *defined* in terms of the numeric computational details, $J^G \equiv J^C$. Thus, since direct search methods require only J^C — and although possibly performance-inferior to idealized derivative-based methods in several ways — they are very easy to use and typically decently successful. These include the Nelder-Mead simplex C_{NM} , and the genetically inspired CMA-ES, $C_{\text{CMA-ES}}$, and differential evolution, C_{DE} .

We have seen, however, that correctly aligning the computational derivatives $\nabla J^C = \nabla J^G$ and $\nabla^2 J^C = \nabla^2 J^G$ required by local optimizers is much harder and typically more expensive. This is because a successful C must take into account implementation details responsible for G such as time-evolution operator, problem representation, the control parametrization and so on. In Sec. 3.3 we discussed three distinct candidates: analytical derivatives C_{ana} , automatic differentiation C_{auto} , and numerical finite-differencing C_{num} . There, C_{ana} was

highlighted as preferable because it is machine-precision-exact to G as is C_{auto} but faster to evaluate, and C_{num} is not exact to begin with. Although we did not discuss it in detail at the time, we have also seen yet another computation that is machine-precision-exact to G in Figs. 4.5, 4.7, namely the auxiliary matrix method [3, 263, 277], C_{aux} . There, one utilizes the relation

$$\begin{pmatrix} \hat{\mathcal{U}}_n & \frac{\partial \hat{\mathcal{U}}_n}{\partial u_n} \\ 0 & \hat{\mathcal{U}}_n \end{pmatrix} = \exp \left[-i \begin{pmatrix} \hat{H}_n & \hat{H}_n^{c'} \\ 0 & \hat{H}_n \end{pmatrix} \delta t \right] \quad (4.76)$$

by computing the right-hand 2×2 block exponential and then extracting the propagator and gradient from the resulting matrix (this exponential can be augmented to a 3×3 block exponentiation to also include the Hessian). Another possible computation relies on diagonalizing the Hamiltonian at each time step [7], $C_{\text{diagonalize}}$. Among these analytical exact procedures, the control diagonal Trotter version of C_{ana} presented here and in Ref. [3] is, to my knowledge, the only one not to be exponentially hampered by Hilbert space dimension when machine-precision-exactness $C = G$ is required. The price paid by using C_{ana} lies in the initial analytical derivation and programming the equations in Sec. 4.4.

Finally, upon choosing an appropriate computational scheme C which hopefully coincides with G , we can then produce seeds according to some strategy and invoke the nonlinear optimization machinery of Sec. 4.1 (search direction, line searching, etc.) and traverse the control landscape.

4.6.2 Of Algorithms and Acronyms

Depending on one's familiarity with the field, the exposition above may or may not seem overly explanatory. Nevertheless, one of its main intended purposes is to clearly delineate what is “quantum” and what is standard machinery from optimization literature. Concretely, *any* computational scheme C invariable contains the “quantum-part” — i.e. dynamics, Hamiltonian, etc. — wherefore it is also often referred to as a *quantum optimal control algorithm* (henceforth simply *algorithm*). Because of their special significance, a typical situation is then that these algorithms are designated by illustrious and/or delightfully witty acronyms such as GRAPE [59], CRAB [241], GROUP [165], GOAT [223], and so on. Overall, this is GREAT¹⁶. Concretely, the acronyms often end up circumscribing more than just C , for example the search direction, time-evolution approximant, and so on. This can lead to a potentially confusing and ambiguous terminology. Especially, methodological developments often formulate “extensions” of an algorithm. It is not that these are not valuable, rather just that continued acronym

¹⁶ Generally Reasonably Effective Albeit slightly Touchy

labeling and re-labeling is not necessarily, in my opinion, all that helpful. It is emphasized that this is mainly a personal observation rather than a critique — of course succinct nomenclature it not without merit and has without a doubt helped standardize and advance the field, even if it incurs some pitfalls here and there. We could name our own “extension” ST-CDR (Suzuki Trotter with Control Diagonal Representation).

It is appropriate to compare the material presented here to the named standard algorithms. These are implemented across many different software libraries, e.g. DYNAMO [283], SPINACH [284], KROTOV [224], QENGINE [1], and QUTIP [285], just to name a few.

Gradient Ascent Pulse Engineering (GRAPE) — One of the most famous and cited derivative-based algorithms is GRAPE [59]. Its original formulation is identical to the discretize-then-optimize first order approximation of Eq. 4.73b in a slightly different notation. This name has stuck since then and now there are many successful variations (“extensions”) on this same theme [7, 263, 271] and is generally associated with full resolution piecewise constant control parameters. Exact gradients are calculated in e.g. Ref. [271] which also utilizes a recursive commutator approach — i.e. including the infinite tail beyond first order — and in this sense the material presented in Sec. 4.4 qualifies as another “extension¹⁷” hereto where the propagator and representation details are accounted for and the exact Hessian is calculated.

It is remarked that first-order results similar to first-order GRAPE were obtained two decades prior [286] using a optimize-then-discretize variational derivations akin to that in Sec. 4.3.

Gradient Optimization Using Parametrization (GROUP) — An algorithm for treating derivative-based optimization in an arbitrary parametrized control basis (i.e. not just piecewise constant) is GROUP. It was derived [165] variationally and yields Eq. (4.43). Section 4.4 derives equivalent expressions in the discretize-then-optimize setting and could then also qualify as extension hereof¹⁸. The chain rule for variations and usual derivatives in fact yield identical equations except that the summation is over the respective optimize-then-discretize and discretize-then-optimize derivatives.

Gradient Optimization of Analytic Controls (GOAT) — The GOAT algorithm calculates the gradient for bilinear Hamiltonians where the control coefficients, call them $(u)_n = u_n$, are analytically parametrized in a function basis (similarly to GROUP). It is derived [223] using

¹⁷ “GRAPE-ST-CDR-BFGS”, perhaps, but I doubt it will catch on.

¹⁸ “GROUP-ST-CDR-BFGS”.

a neat trick yielding coupled equations for the propagator and its derivative,

$$\frac{\partial}{\partial t} \begin{pmatrix} \frac{\partial \hat{U}}{\partial u} \\ \hat{U} \end{pmatrix} = -\frac{i}{\hbar} \begin{pmatrix} \hat{H} & \frac{\partial \hat{H}}{\partial u} \\ 0 & \hat{H} \end{pmatrix} \begin{pmatrix} \frac{\partial \hat{U}}{\partial u} \\ \hat{U} \end{pmatrix}, \quad \hat{U}(0) = \mathbf{1}, \quad \frac{\partial \hat{U}(0)}{\partial u} = \mathbf{0}, \quad (4.77)$$

paraphrased here to fit notation. This represents a generally large system of coupled equations to be solved from $t = 0 \rightarrow T$, one equation for each control parameter and for both the propagator and derivative matrices (individually of size $D_{\mathcal{H}}^2$). As far as I can tell, it is an independent discovery of equations formally equivalent to the auxiliary matrix method [263, 277] (compare Eq. 4.76). One difference, though, is in their suggested solution where Ref. [223] proposes ODE solvers. Although the second caution of Sec. 4.5.2 no longer applicable in this regard, the first and third remain.

Krotov— Another popular and proven class of derivative-based algorithms are the KROTOV variants [224, 274, 287–289]. Originally due to V. F. Krotov, these are genuinely worth their own label because they are disambiguous in applying a quite distinct trick. Briefly, it involves adding to the original cost J an auxiliary, vanishing term (i.e. the landscape is unchanged). This term can be chosen in such a way that variational derivatives yield update equations for individual control points u_n that monotonically decreases J (only strictly ensured in the continuous time limit).

The monotonic decrease property is typically highlighted, likely because of the very clever trick utilized in ensuring it, although it should be noted that most optimization methods has this feature since the search direction is nearly always a descent direction — if a method does *not* have this property it is typically by construction, e.g. in allowing a sub-optimal choice in this iteration for a potentially larger reward in the following iterations.

The algorithm is associated with full resolution piecewise constant parametrization and KROTOV updates individual control points in sequence and is therefore aptly named a *sequential* algorithm. It typically does so with a constant step size length i.e. without line searching. When KROTOV is compared with GRAPE the latter is occasionally [283, 290] named a *concurrent* algorithm because it (typically) updates all points at the same time. I do not think, however, that this is entirely accurate because “GRAPE-like” algorithms can also just as easily update individual control points sequentially and in standard optimization literature this is called coordinate ascent/descent [214].

Neither of these two somewhat semantic observations are of detriment to either KROTOV nor GRAPE algorithms.

Chopped Random Basis (CRAB) — Another one of the most widely applied *methodologies* is CRAB. As introduced in Sec. 3.5 this name pertains to a particular parametrization of the control, namely into a chopped random basis (a complete basis truncated to be finite with random elements). Thus, CRAB is strictly speaking not really an algorithm nor an optimizer and was not presented as such originally [73, 241]. Yet, through widespread adoption, the acronym CRAB has become more or less synonymous with derivative-free Nelder-Mead simplex updates [242, 290–292] — despite the fact that any direct search method such as CMA-ES and DE could be just as easily utilized. However, derivative-based methods such as GROUP and GOAT could also be considered a part of CRAB since they, too, entail on a functional basis expansion. Further, if the expansion is not *decidedly* onto a set of N functions that becomes a complete basis as $N \rightarrow \infty$ then that expansion cannot really be considered a “chopped basis” and hence such parametrizations are not a strictly a CRAB despite covering basically the same intent.

The CRAB “algorithm” — which is not really an algorithm — is perhaps therefore a good example of a somewhat confused and ambiguous terminology. Further, one might also argue that GROUP is “just” GRAPE using the chain-rule and then, in a certain sense, CRAB, GRAPE, GOAT, and GROUP are all shades of the same thing because of their partially overlapping definitions. This is again quite semantic without any actual detriment to the optimization itself, but it illustrates the source for potential confusion.

This concludes our summary and perspectives of quantum optimal control algorithms. It is noted that the variational ones can also be formulated using the Pontryagin maximum principle [215, 272, 293]. Another derivative-free approach, Stochastic Ascent (SA), is described and utilized in Ref. [5].

Probabilistic optimal control such as Bayesian approaches [294] are considered outside the scope of this thesis.

MANUSCRIPT [1] – QENGINE: A C++ LIBRARY FOR QUANTUM OPTIMAL CONTROL OF ULTRACOLD ATOMS

Published in Computer Physics Communications.
16 pages.

AUTHOR CONTRIBUTIONS

In the development phase, Jens Jakob Sørensen was the project lead. The software design and planning was a joint effort between mainly Jens Jakob Sørensen, Till Heinzl, and myself. Till Heinzl was the lead programmer. I also contributed to parts of the source code and wrote the example programs as well as many of the unit tests, and I was the code reviewer on merge requests.

In the writing phase, Jens Jakob Sørensen and I wrote the paper with equal contribution to the writing process with scientific feedback and corrections from Jacob Sherson.

MANUSCRIPT [3] – APPROXIMATE DYNAMICS LEAD TO MORE OPTIMAL CONTROL: EFFICIENT CALCULATION OF EXACT DERIVATIVES

*In review with Physical Review A. Favorable report received.
16 pages.*

AUTHOR CONTRIBUTIONS

Some of the exact control derivative calculations (those of the $\hat{U}^{\text{ST}_1}$ and $\hat{U}^{\text{Ex}}$ propagators in the manuscript) were initially performed by Frederik Møller (during his master's project) and Jens Jakob Sørensen. At that time, there was no delineation between this project and what has since evolved into manuscript [4].

When I continued the project, I streamlined, formalized, and fixed details of the existing calculations. I introduced the interpretation in terms of the control landscape, i.e. the difference between being exact with respect to the dynamics versus in the derivative. I calculated the exact control derivatives for the $\hat{U}^{\text{ST}_2}$, for the other standard propagators (Chebyshev, Crank-Nicolson, and Krylov-Lanczos, not shown explicitly in here or the manuscript), and for regularization terms. I extended the results to cover all unitary problems with state-transfer being a special case. I performed the generalizability analysis to multiple controls.

I wrote the numerical software from scratch (distinct from the code used in manuscript [4]), and performed the numerical studies, comparisons, data analyses, and interpretations.

It became clear that the significantly increased scope warranted two separate publications, this and manuscript [4].

In the writing phase, I authored the manuscript and figures with scientific feedback and corrections from Frederik Møller and Jacob Sherson.

BOSE-EINSTEIN CONDENSATE PREPARATION WITH EXACT GRADIENTS

AUTHOR CONTRIBUTIONS

I performed the analytical and numerical studies and implementations, analysis and interpretation of the presented results, and authored the text and figures.

THIS short chapter details an, as of writing, unpublished work in progress related to arbitrary state transfers of meanfield [BEC](#)—nonlinear Schrödinger equations with nonzero self-interactions—using analytically exact gradients. The corresponding Hessian derivations are left for future work. Several equations and concepts are restated and some symbols are redefined without particular mention for self-containedness.

The chapter is structured as follows.

Section [5.1](#) sets the numerical stage and points out difficulties in and solutions to diagonalizing the stationary Gross-Pitaveskii equation ([GPE](#)) beyond the ground ($i = 0$) and first excited state ($i = 1$).

Section [5.2](#) exemplifies a state transfer problem involving the second excited [GPE](#) state ($i = 2$) and highlights potential research questions and preliminary results using optimize-then-discretize gradients.

Section [5.3](#) extends the discretize-then-optimize results of [Sec. 4.4](#) by calculating analytically exact gradients for arbitrary self-interaction strengths. These are numerically verified and it is similarly numerically shown that the corresponding optimize-then-discretize gradients are, roughly, first-order approximations hereto.

5.1 NUMERICS OF MEANFIELD BEC

The [GP](#) Hamiltonian in the $\hat{x}$ -representation reads

$$\hat{H}_{\text{gp}}(x, u(t); \psi, G_{1\text{D}}) = \hat{T} + \hat{V}(x, u(t)) + G_{1\text{D}}|\psi(x)|^2, \quad (5.1)$$

with the kinetic energy $\hat{T} = -(\hbar^2/2m)\partial^2/\partial x^2$, the potential $\hat{V}$ controlled with $u(t)$, and the [BEC](#) self-interaction strength which is generally a function of the condensed number of atoms, $G_{1\text{D}} = G_{1\text{D}}(N_{\text{p}})$. For specificity we consider repulsive interactions $G_{1\text{D}} \geq 0$ in this section although this does not have any impact on the numerics or

algorithms. Likewise, we consider (quasi-)one-dimensional geometries but the presented calculations hold on in all dimensions.

Numerically solving the time-dependent GPE with exact exponentiation $\exp(-i\hat{H}_{\text{gp}}(\psi)\delta t/\hbar)|\psi\rangle$ is, as in the single-particle case equivalent to $G_{1\text{D}} = 0$, not an effective approach: its runtime complexity scales as $\mathcal{O}(N_x^3)$ with Hilbert space sizes typically of order $N_x \sim 10^2$ induced by the numerical discretization $x \in [x_1, \dots, x_{N_x}] = [x_{\min}, \dots, x_{\max}]$ with equidistant spacing δx . Fortunately, a fairly simple modification (specified later in Sec. 5.3.2) to the single-particle split-step Fourier Trotterization scheme [264] admits a $\mathcal{O}(N_x \log N_x)$ scaling when using fast Fourier transforms with the same low error of order $\mathcal{O}(\delta t^3)$. This extension is possible since the self-interaction term $G_{1\text{D}}|\psi(x)|^2$ is diagonal in $\hat{x}$ and basically acts as an additional potential.

Numerically solving, however, the general time-independent GPE,

$$\hat{H}_{\text{gp}}(\phi_i, G_{1\text{D}})|\phi_i\rangle = E_i|\phi_i\rangle, \quad (5.2)$$

for an arbitrary eigenpair $\{E_i, |\phi_i\rangle; G_{1\text{D}}\}$ is non-trivial. This is because the nonlinearity of Eqs. (5.1)–(5.2) generally break [295] the usual orthogonality conditions¹ $\langle\phi_i|\phi_j\rangle \neq \delta_{i,j}$ and thus also, e.g., the usual approach of projecting out prior states $\{|\phi_0\rangle, \dots, |\phi_{i-1}\rangle\}$ during a state energy variance functional minimization. Note that solving for the ground state $i = 0$ — and possibly the first excited state $i = 1$ if $|\phi_{0,1}\rangle$ are of opposite parity — can still be done quite easily by the usual imaginary time-evolution, and a number of quantum control problems concerns only transfers between these [162, 167, 197]. For general, arbitrary BEC state transfer problems, however, more advanced diagonalization algorithms must be considered such as the one detailed in Ref. [295]. The results presented in this section are based on our implementation of this. A full account of the algorithm is not particularly stimulating, but the general gist is the following. For some i , the interaction-free ($G_{1\text{D}}^{(0)} = 0$) eigenpair $\{E_i^{(0)}, |\phi_i^{(0)}\rangle; G_{1\text{D}}^{(0)}\}$ is calculated by any usual diagonalization means where the superscript denotes an iteration index. Then, the right hand side of Eq. (5.2) is increased by a sufficiently small energy, $E_i^{(1)} = E_i^{(0)} + \delta E$. Taking $|\phi_i^{(0)}\rangle$ as an initial guess, a derivative-based minimization² is then performed over an appropriate functional for $|\phi_i^{(1)}\rangle$ which corresponds to an indirect, nonlinear increase to $G_{1\text{D}}^{(1)} = G_{1\text{D}}^{(0)} + \delta G_{1\text{D}}$. From the new eigenpair, $\{E_i^{(1)}, |\phi_i^{(1)}\rangle; G_{1\text{D}}^{(1)}\}$, the process can be repeated as necessary up to iteration k by increasing the energy $E_i^{(k)} = E_i^{(k-1)} + \delta E$ (the chosen δE can of course vary between iterations). The output is thus sets of eigenpairs $\{E_i^{(j)}, |\phi_i^{(j)}\rangle; G_{1\text{D}}^{(j)}\}_{j=1}^k$ of increasing interaction strengths, but same quantum mode i (wave function modality).

¹ It is numerically customary to explicitly factor out the number of atoms from the condensate wave function $\psi \rightarrow \sqrt{N}\psi$ such that ψ is unit normalized.

² The necessary equations are formed by utilizing the Gâteaux variations of Sec. 4.3.

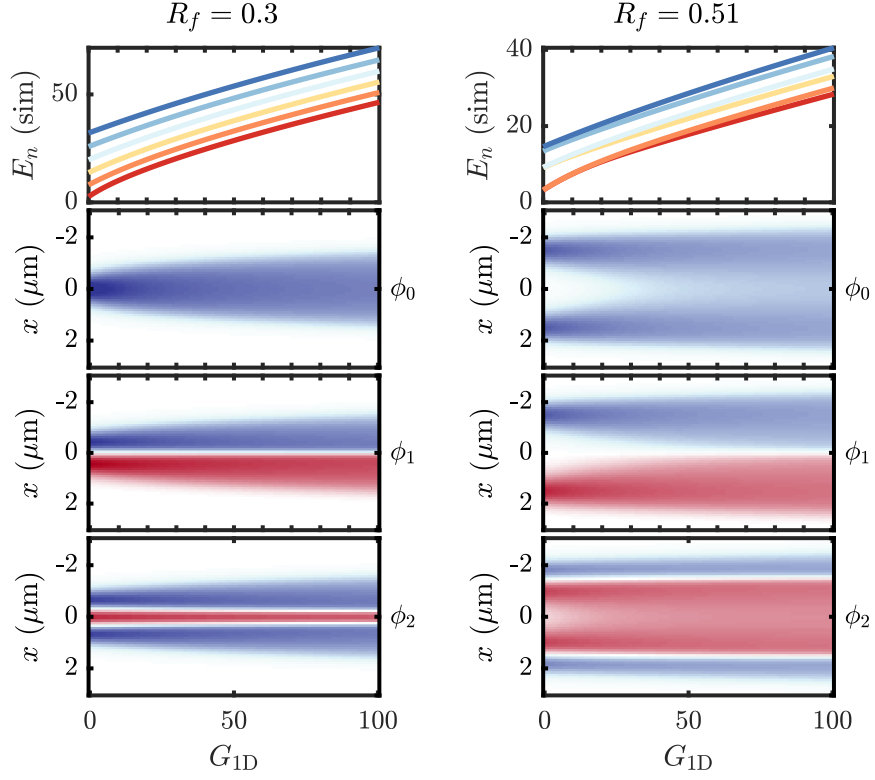


Figure 5.1: Energy spectrum associated with atom chip potential $V(R_f)$ in Eq. (5.3) as a function of G_{1D} . Here, $R_f = 0.3$ (left) and $R_f = 0.51$ (right) correspond to anharmonic single and double wells, respectively. The ϕ_i color maps indicate positive (negative) values in blue (red) and white is zero.

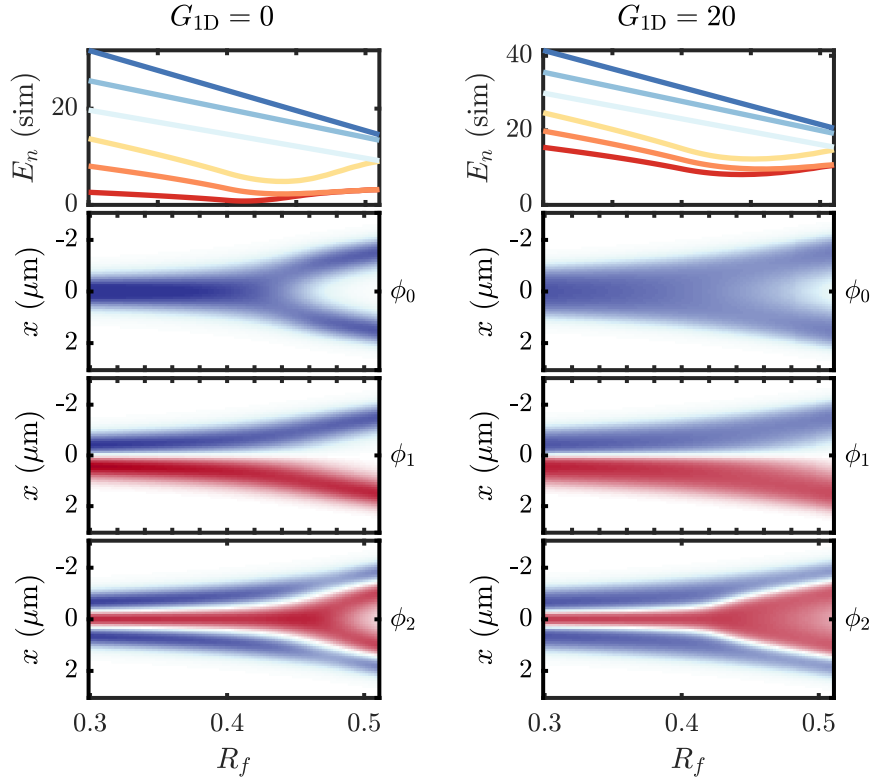


Figure 5.2: As Fig. 5.1, but for fixed $G_{1D} = 0$ (left) and $G_{1D} = 20$ (right) and variable R_f showing the single-to-double-well transformation.

In practice, however, one is more interested in obtaining eigenstates corresponding to specified G_{1D} (or number of condensed atoms N_p) rather than the actual energies. In our implementation this is achieved simply by performing a Nelder-Mead simplex search over the objective $\min_{E^{(k)}} |G_{1D}^{\text{target}} - G_{1D}^{(k)}|$ (or similar for N_p). As a rule of thumb, it seems that incremental values of $\delta G_{1D} \leq 10$ typically ensures proper convergence. Additionally, although argued against on principle in Ref. [295], we have found that implementing a linesearch on τ in Step 6 works remarkably well in practice and reduces the algorithm to only a few iterations per eigenpair (down from sometimes hundreds). With this algorithm, energy eigenpairs corresponding to arbitrary self-interactions strengths or atom numbers can be determined to machine precision, i.e. $\text{var}(\hat{H}_{\text{gp}})_{\phi_i} = \langle \phi_i | \hat{H}_{\text{gp}}^2 | \phi_i \rangle - \langle \phi_i | \hat{H}_{\text{gp}} | \phi_i \rangle^2 \lesssim \epsilon_{\text{mach}}$ where $\epsilon_{\text{mach}} \sim 10^{-16}$.

As example, an energy spectrum for $i = 0, 1, \dots, 5$ and associated wave functions for $i = 0, 1, 2$ are shown in Fig. 5.1–5.2. The system and potential parameters are described in the following section.

5.2 DOUBLE TWIN BEAM SOURCE INITIALIZATION

One example of a state transfer problem that involves excited states $i > 1$ is the recently reported rubidium 87 atom chip BEC experiment [166, 296]. There, optimal control was used to drive the ground state ($i = 0$) of an anharmonic single-well potential configuration into the second excited state ($i = 2$) of an anharmonic double-well configuration which acts as a source for double twin atom beams. Briefly, atoms in the source state can undergo pairwise collisions and be ejected along either well — with opposite momenta due to conservation laws — which upon recombination results in a measurable two-particle interference pattern [166, 296].

In a certain sense it is a combination of the *Shake Up* and *Splitting* problems described in Ch. 8 originating from the same experimental group and atom chip. As in those cases, a quasi-one-dimensional treatment is approximately valid (cf. Sec. 2.4) at sufficiently short time scales as follows because of the trapping frequency hierarchy $\omega_z \ll \omega_x < \omega_y$. Dynamics in the axial (ejection) z direction occurs on much longer time scales than in the transverse x and y directions and assuming that the potential is approximately separable $\hat{V}(x, y, z) \approx \hat{V}^{(x)}(x) + \hat{V}^{(y)}(y) + \hat{V}^{(z)}(z)$, the y direction is frozen out³. This allows an approximate decomposition into the product form $|\psi^{(x)}(t)\rangle |\phi_0^{(y)}\rangle |\phi_0^{(z)}\rangle$ and this justifies considering only $|\psi(x, t)\rangle = |\psi^{(x)}(t)\rangle$ and $\hat{V}(x) = \hat{V}^{(x)}(x)$ in the numerical treatment (the state remains in the ground state $|\phi_0^{(y)}\rangle |\phi_0^{(z)}\rangle$ along y and z). Note that this modeling inherently does not include the collisional processes

³ Note that we have swapped the x and y labels compared to Ref. [166].

leading to the double twin beams (i.e. along the z direction). For the single twin atom beam problem (*Shake Up*) this occurs on time scales of ≈ 3 ms and in both instances no known intuitive, e.g. adiabatic, solutions are known. Protocols facilitating this type of transfer must therefore be well below this time to maximize measurement purity and, incidentally, this also points to the need for quantum optimal control and driving near the minimal duration $T_{\min}^{F=1}$.

Denoting the control as $R_f(t)$, the control potential can be written as a power expansion

$$\hat{V}(x, R_f(t)) = \sum_{i=0,2,4,6} a_i(R_f(t))x^i, \quad (5.3)$$

and is shown in Fig. 5.3. This has a similar form as the potential Eq. (2.10) except the control is now in the coefficient a_i . These are given in Ref. [166] as sixth order polynomials in R_f , calibrated by a fitting procedure between numerically simulated and experimental data. The Hamiltonian is then $\hat{H}_{\text{gp}}(R_f, G_{1\text{D}}) = \hat{T} + V(R_f) + G_{1\text{D}}|\psi|^2$ and the initial state is taken to be $|\psi_{\text{ini}}\rangle = |\phi_0; R_f = 0.3, G_{1\text{D}}\rangle$ and the target state to be $|\psi_{\text{tgt}}\rangle = |\phi_2; R_f = 0.51, G_{1\text{D}}\rangle$ for some value of the quasi-one-dimensional self-interaction strength $G_{1\text{D}}$. In this geometry it can be written $G_{1\text{D}}(N_p) = g_{1\text{D}}(N_p)N_p$ where $g_{1\text{D}}(N_p)$ is calculated as in Ref. [74]. The result is shown in Fig. 5.4. To calculate the $i = 0, 2$ eigenstates of the nonlinear GPE we use the algorithm described in the previous section⁴. Figures 5.1–5.2 shows the energy spectrum as a function of $G_{1\text{D}}$ with fixed R_f and vice versa, respectively, verified to have an energy variance of zero to machine precision.

For fixed R_f in Fig. 5.1 increasing the repulsive interaction $G_{1\text{D}}$ flattens and broadens ϕ_i as expected. In particular, the ground states become increasingly Thomas-Fermi-like [158], $|\phi_0| \sim \sqrt{(E_0 - V(x))/G_{1\text{D}}}$ (compare Fig. 5.3). For $R_f = 0.51$ this means that the initially well-localized population is pushed into the (classically forbidden) barrier region. The energies remain slightly anharmonic for $R_f = 0.3$ whereas the interaction term is seen to lift the pair-wise symmetric/anti-symmetric degeneracy for $R_f = 0.51$.

For fixed $G_{1\text{D}}$ in Fig. 5.2 — depending on the value — the spectrum goes from being nearly harmonic to being nearly pair-wise symmetric/anti-symmetric degenerate and the states become increasingly broadened.

It is remarked, as in Fig. 5.4, that a typically cited number of BEC atoms in these magnetic traps is $N_p = 700$ yielding $G_{1\text{D}}(700) = 2.5$.

⁴ The energies and durations are presented here in simulation units. It seems to me that the numbers stated in the supplementary of Ref. [166] are not quite fully consistent (they are stated half in SI units and half in non-dimensionalized units). After some modifications we obtain sensible numerical results, e.g. the correct well configurations for $R_f = 0.3, 0.51$ and $\text{var}(\hat{x})_{|\phi_i; R_f=0.3\rangle} = 0.161 \mu\text{m}$ which is roughly consistent with $r_0 = 0.172 \mu\text{m}$.

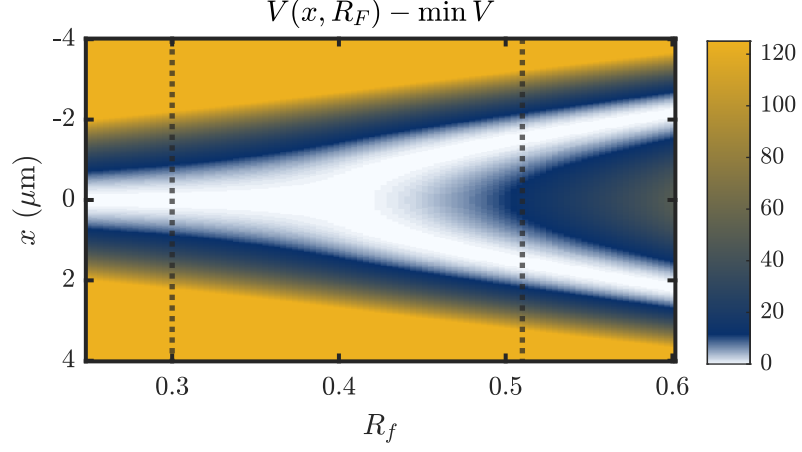


Figure 5.3: Double twin beam potential as a function of control value R_f . The dotted lines indicate the values defining single-well (double-well) configuration for the initial state $|\psi_{\text{ini}}\rangle = |\phi_0; R_f = 0.3, G_{1D}\rangle$ (target state $|\psi_{\text{tgt}}\rangle = |\phi_2; R_f = 0.51, G_{1D}\rangle$). The color scale is chosen to highlight the well structure and clips very large values $V \geq 120$.

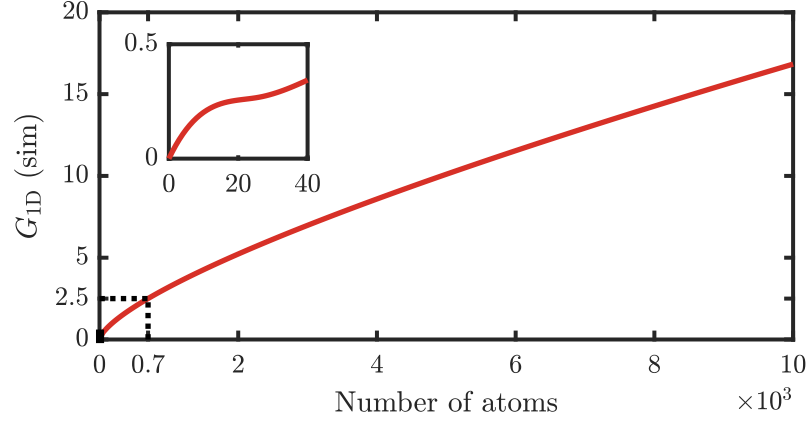


Figure 5.4: $G_{1D}(N_p) = g_{1D}(N_p)N_p$ as a function of condensed atoms. The highlighted $G_{1D}(700) = 2.5$ corresponds to the number of condensate atoms used in Refs. [5, 74, 166]. The inset highlights behavior in low the N_p regime.

This is obviously very far from e.g. the $100G_{1D}$ shown in the spectra. The reason for extending axis to such high interaction values is twofold. First, it nicely illustrates the exaggerated interaction effects and second, *optical* traps routinely trap millions of atoms where these would not be unrealistic. This points to the broader fact that the considered problem is chosen on the basis of recency and concerning a $i = 0 \rightarrow 2$ transfer — the following research questions could be studied exactly the same for any other transfer and trapping geometry. This includes optical traps where it seems reasonable to also expect much larger *fluctuations* in atom number, equivalently interaction strength, and the questions are thus likely even more experimentally relevant there.

Research questions

It is quite clear from the previous section that the initial and target states depend on the interaction strength G_{1D} , correspondingly atom number N_p . Among others, this leads to a few possible directions of study, for example in the double twin beam setup described above or in other related geometries:

- Q1** — Mapping out the minimal durations to reach a certain fidelity threshold, $T_{\min}^F(G_{1D})$, do these depend strongly on G_{1D} ?
- Q2** — What are the optimal control strategies? Are there many, few, or just a single one? Do these depend on G_{1D} and/or T ?
- Q3** — Are the solutions stable to fluctuations in atom number and thus G_{1D} ? If not, can stability be achieved by ensemble optimization [62]?

Preliminary results

In Fig. 5.5 we very briefly present a single batch of optimization results at different G_{1D} and T . The optimizations were performed with the QENGINE implementation of the optimize-then-discretize Eqs. (4.38) ("GRAPE") using the BFGS direction with line searching. The control was constrained as $0.25 \leq R_f(t) \leq 0.6$ using soft-bounds of strength $\sigma = 2 \cdot 10^3$ and derivative regularized with strength $\gamma = 10^{-8}$. For each combination of G_{1D} and T , 200 seeds were optimized. The seeds were produced by first generating a reference control consisting of three random straight line segments and overlaying these with Fourier terms (between 5 and 40 components of increasing harmonics) with randomized strengths, phases, perturbative frequencies, and overall fluence.

We achieve a rough estimate for a minimal duration, $T_{\min}^{F=0.99} = 1.4$, and find indications from the histograms that different G_{1D} might play a role in the abundance of solutions (Q1) — this can nevertheless not be substantially concluded on the basis of this relatively sparse,

first-run dataset. However, looking at the high-fidelity solutions of Fig. 5.6 we find indications of both prominent bang-bang and smoother oscillatory strategies (Q2). The $F = 0.99$ dynamics implemented by the highlighted optimized bang-bang control are shown in Fig. 5.7. It is mentioned in passing that around $T = 0.6$, the condensate density closely resembles that of the single-well configuration second excited state, $|\phi_2; R_f = 0.3\rangle$, i.e. $\int_{-\infty}^{\infty} |\psi| |\phi_2| dx \approx 0.99$. At the same time, the corresponding fidelity is yet only 0.42 because of the phase associated with the motion of ψ .

From the optimization trajectories in Fig. 5.5 we find that many solutions flatten out rather quickly. This can simply be due to converging to an optimum. However, since we show in the following section that the gradients employed in this optimization are in fact not exact, another plausible explanation is that the BFGS approximation is scrambled and the convergence is false just like in Fig. 4.7. Thus, as described in Sec. 4.5, even though the ensemble robustness (Q3) is not yet implemented here it would likely be extensively hampered by these inaccuracies. In general it is guaranteed that using the exact gradient will yield better results than those presented here, granted they are not exponentially expensive to calculate — which they indeed turn out not to be.

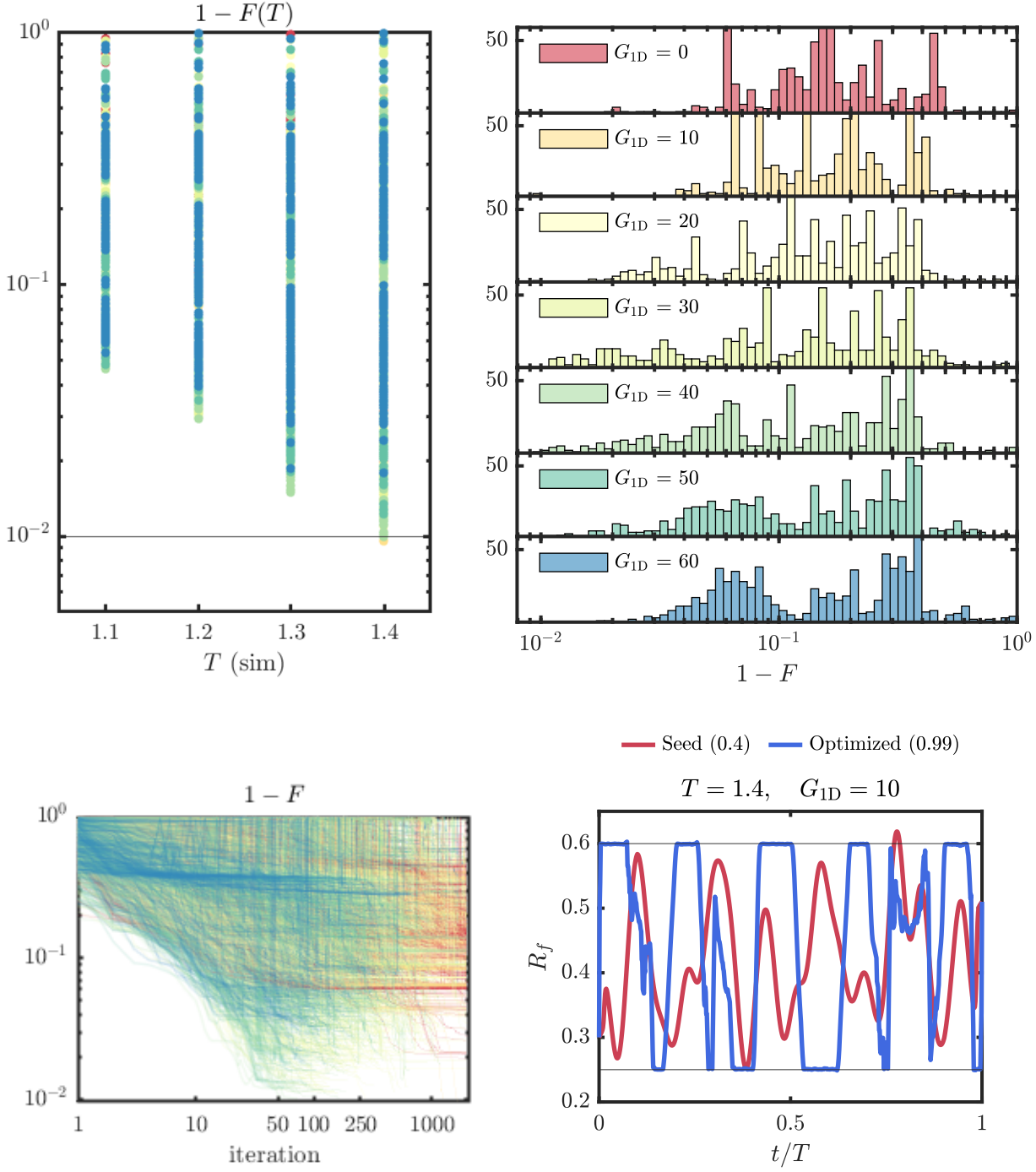


Figure 5.5: Preliminary double twin beam results. Optimized solutions shown in $F(T)$ space (top left), in histograms sorted by interaction strength (top right), and optimization trajectories (bottom left). An example seed ($F = 0.4$) and optimized solution ($F = 0.99$) shows a bang-bang-like structure (bottom right).

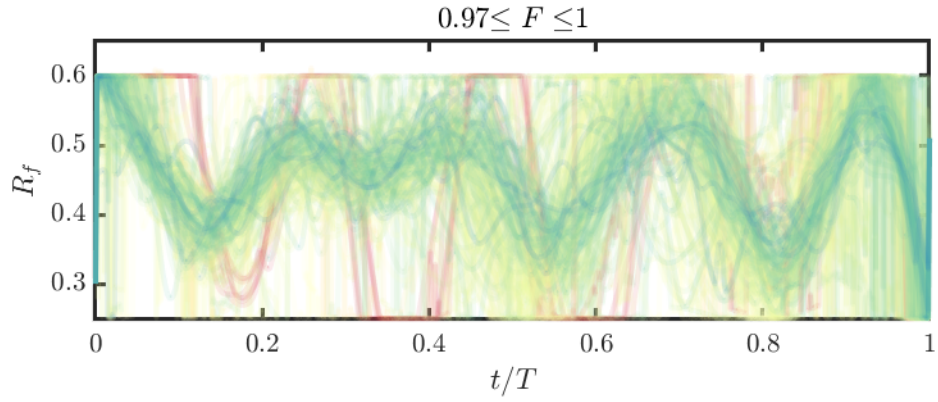


Figure 5.6: High-fidelity solutions. These results suggest that there are both bang-bang and oscillatory strategies present in the control landscape.

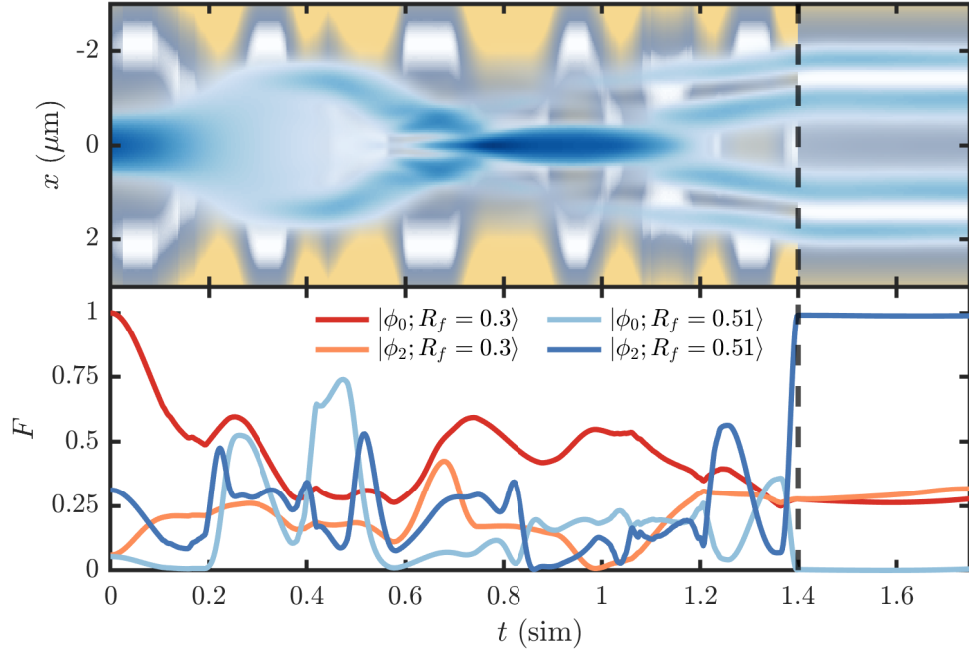


Figure 5.7: Optimal BEC bang-bang dynamics for driving $|\phi_0; R_f = 0.3\rangle \rightarrow |\phi_2; R_f = 0.51\rangle$. The control $R_f(t)$ was optimized up until the vertical line at $T = 1.4$ and is then held static at $R_f = 0.51$. Top: Condensate density $|\psi(x, t)|^2$ (blue colormap) and atom chip potential $V(x, R_f(t))$ (same colormap as Fig. 5.3 but slightly translucent). Bottom: Fidelities with different even energy eigenstates (odds are not coupled because of symmetry). The desired transfer is achieved with a fidelity of 0.99.

5.3 EXACT ANALYTICAL GRADIENTS FOR MEANFIELD BEC

Here we calculate exact analytical gradients using discretize-then-optimize as in Sec. 4.4 for a BEC meanfield with $G_{1D} \neq 0$. Emphasis is put on thoroughness of the steps and relevant equations for the derivations are restated for convenience where applicable so as to be self-contained.

The state transfer fidelity cost and its gradient can be written

$$J_F = \frac{1}{2}(1 - F(T)) = \frac{1}{2}(1 - o^* o), \quad (5.4a)$$

$$\nabla J_F = -\Re(o^* \nabla o), \quad (5.4b)$$

$$o = \langle \chi_{N_t} | \psi_{N_t} \rangle = \langle \chi_{N_t} | \hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_n \dots \hat{\mathcal{U}}_1 | \psi_1 \rangle. \quad (5.4c)$$

Here we defined the initial state $|\psi_1\rangle = |\psi_{\text{ini}}\rangle$ and auxiliary state $|\chi_{N_t}\rangle = |\psi_{\text{tgt}}\rangle$ on the usual discretized time grid of regular δt intervals, leading to piece-wise constant time evolution and Hamiltonian operators

$$t \in [t_1, t_2, \dots, t_{N_t}] = [0, \delta t, \dots, T], \quad (5.5a)$$

$$\hat{\mathcal{U}}(T; 0) = \hat{\mathcal{U}} = \prod_{j=1}^{N_t-1} \hat{\mathcal{U}}_j = \hat{\mathcal{U}}_{1:N_t-1} = \hat{\mathcal{U}}_{N_t-1} \dots \hat{\mathcal{U}}_2 \hat{\mathcal{U}}_1, \quad (5.5b)$$

with time indices denoted as subscripts and T is the total duration ($N_t = \lfloor T/\delta t \rfloor + 1$). The states are progressed as usual, e.g. $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n |\psi_n\rangle$ and $|\chi_n\rangle = \hat{\mathcal{U}}_n^\dagger |\chi_{n+1}\rangle$, with the added complication that now $\hat{\mathcal{U}}_n$ depends on the ψ state history as described in more detail below.

5.3.1 Piecewise Exact Propagator

Let $\hat{V}_n = \hat{V}(x, u_n)$ be the potential with control parameter $u_n = u(t_n)$ and define $\alpha = -i\delta t/2$. The time-evolution operator at time index n , $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n |\psi_n\rangle$, explicitly depends on all prior operators and thus control values $u_{1:n}$,

$$\hat{\mathcal{U}}_n = \exp(2\alpha \hat{H}_n), \quad (5.6)$$

$$\hat{H}_n = \hat{T} + (\hat{V}_n + G_{1D}|\psi_n|^2) = \hat{T} + (\hat{V}_n + G_{1D}|\hat{\mathcal{U}}_{1:n-1}\psi_1|^2), \quad (5.7)$$

in units where $\hbar = 1$. That is, $\hat{\mathcal{U}}_n = \hat{\mathcal{U}}_n(\{u_{1:n}\})$, or, in other words, the control u_n affects all following time-evolution operators $\hat{\mathcal{U}}_{n:N_t-1}$. This fact quite complicates the derivative calculations below. We remark further that

$$\psi_n = \langle x | \psi_n \rangle = \langle x | \hat{\mathcal{U}}_{n-1} | \psi_{n-1} \rangle = \hat{\mathcal{U}}_{n-1} \langle x | \psi_{n-1} \rangle = \hat{\mathcal{U}}_{n-1} \psi_{n-1}, \quad (5.8)$$

$$\begin{aligned} |\psi_n|^2 &= \psi_n^* \odot \psi_n = \psi_n^* \odot (\hat{\mathcal{U}}_n \psi_{n-1}), \\ &\neq \psi_n^* \hat{\mathcal{U}}_n \odot \psi_{n-1} = \psi_{n-1}^* \odot \psi_{n-1} = |\psi_{n-1}|^2, \end{aligned} \quad (5.9)$$

where $\odot$ denotes element-wise (Hadamard) multiplication. It is quite obvious that generally $|\psi_n|^2 \neq |\psi_{n-1}|^2$, especially when written out like this. Nevertheless, the $\odot$ symbol is usually not explicitly shown and this can lead to a host of mistakes and faulty conclusions that are easy to make but hard to identify. Thus, it pays to explicitly include $\odot$ when appropriate.

Although this propagator is never used in practice cf. Sec. 5.1, calculating its gradient serves as a good warm up for the Trotterized case which are more involved.

Based on the above, the gradient elements of the transfer amplitude take the form

$$\begin{aligned}
 \frac{\partial o}{\partial u_n} &= \frac{\partial}{\partial u_n} (\langle \chi_{N_t} | \hat{\mathcal{U}} | \psi_1 \rangle) = \left\langle \chi_{N_t} \left| \frac{\partial}{\partial u_n} (\hat{\mathcal{U}}_{1:N_t-1}) \right| \psi_1 \right\rangle \\
 &= \left\langle \chi_{N_t} \left| \frac{\partial}{\partial u_n} (\hat{\mathcal{U}}_{n:N_t-1}) \right| \psi_{n-1} \right\rangle = \left\langle \chi_{N_t} \left| \sum_{q=n}^{N_t-1} \hat{\mathcal{U}}_{q+1,N_t-1} \frac{\partial \hat{\mathcal{U}}_q}{\partial u_n} \hat{\mathcal{U}}_{n:q-1} \right| \psi_{n-1} \right\rangle \\
 &= \sum_{q=n}^{N_t-1} \left\langle \chi_{q+1} \left| \frac{\partial \hat{\mathcal{U}}_q}{\partial u_n} \right| \psi_q \right\rangle \\
 &= 2\alpha \sum_{q=n}^{N_t-1} \left\langle \chi_q \left| \left(\sum_{k=0}^{\infty} \frac{(-2\alpha)^k}{(k+1)!} [\hat{H}_q, \frac{\partial \hat{H}_q}{\partial u_n}]_k \right) \right| \psi_q \right\rangle, \tag{5.10}
 \end{aligned}$$

Note that in the backwards time-evolution of the auxiliary state, $|\chi_q\rangle = \hat{\mathcal{U}}_q^\dagger |\chi_{q+1}\rangle$, the state in the interaction term is $|\psi_q|^2$ and not $|\chi_q|^2$. In the above expression we used a result from Sec. 4.4 to evaluate

$$\frac{\partial \hat{\mathcal{U}}_q}{\partial u_n} = 2\alpha \hat{\mathcal{U}}_q \sum_{k=0}^{\infty} \frac{(-2\alpha)^k}{(k+1)!} [\hat{H}_q, \frac{\partial \hat{H}_q}{\partial u_n}]_k. \tag{5.11}$$

Now, using $(|\kappa|^2)' = (\kappa^* \kappa)' = 2\Re(\kappa^* \kappa')$, we calculate from Eq. (5.7)

$$\frac{\partial \hat{H}_q}{\partial u_n} = \delta_{q,n} \frac{\partial V_n}{\partial u_n} + 2G_{1D} \Re \left(\psi_q^* \odot \frac{\partial}{\partial u_n} \left(\prod_{j=n}^{q-1} \hat{\mathcal{U}}_j \right) \psi_n \right). \tag{5.12}$$

At this point, it is most fruitful to consider this a recursion in

$$\frac{\partial}{\partial u_n} \left(\prod_{j=n}^{q-1} \hat{\mathcal{U}}_j \right) \psi_n = \frac{\partial \hat{\mathcal{U}}_{q-1}}{\partial u_n} \psi_{q-1} + \hat{\mathcal{U}}_{q-1} \frac{\partial}{\partial u_n} \left(\prod_{j=n}^{q-2} \hat{\mathcal{U}}_j \right) \psi_n, \tag{5.13}$$

starting with base case $q = n$:

$$\frac{\partial}{\partial u_n} \left(\prod_{j=n}^{n-1} \hat{\mathcal{U}}_j \right) \psi_n = 0, \quad \frac{\partial \hat{H}_n}{\partial u_n} = \frac{\partial V_n}{\partial u_n}. \tag{5.14}$$

This fully specifies the gradient. It is, however, a numerically very inefficient expression it is a recursion of recursive commutators sums.

5.3.2 Trotterized Control-Diagonal Propagator

As in Sec. 4.4 we would now write $\hat{H}_n = \hat{H}^d + \hat{H}_n^c$. Because we are considering the GPE, however, the (split-step) Trotterization is slightly altered [264] due to the interaction term,

$$|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n^{\text{ST}} |\psi_n\rangle = \hat{\mathcal{U}}_{n+1}^{c++/2} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2} |\psi_n\rangle = \hat{\mathcal{U}}_{n+1}^{c++/2} |\psi_n^{++}\rangle \quad (5.15)$$

with $|\psi_n^{++}\rangle = \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2} |\psi_n\rangle$ and the following propagator and Hamiltonian definitions

$$\hat{\mathcal{U}}_n^{\text{ST}} = \hat{\mathcal{U}}_{n+1}^{c++/2} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2} \quad (5.16a)$$

$$\hat{\mathcal{U}}^d = \exp(2\alpha \hat{H}^d) \equiv \exp(2\alpha \hat{T}), \quad (5.16b)$$

$$\hat{\mathcal{U}}_n^{c/2} = \exp(\alpha \hat{H}_n^c) \equiv \exp(\alpha (\hat{V}_n + G_{1D} |\psi_n|^2)), \quad (5.16c)$$

$$\hat{\mathcal{U}}_{n+1}^{c++/2} = \exp(\alpha \hat{H}_{n+1}^{c++}) \equiv \exp(\alpha (\hat{V}_{n+1} + G_{1D} |\psi_n^{++}|^2)). \quad (5.16d)$$

Note we are considering the ST = ST₁ Trotterization local in n and $n + 1$ controls.

Similar to the exact case we have $\hat{\mathcal{U}}_n^{\text{ST}} = \hat{\mathcal{U}}_n^{\text{ST}}(\{u_{1:n+1}\})$ i.e. the control u_n affects the time-evolution operators $\hat{\mathcal{U}}_{n-1:N_t-1}^{\text{ST}}$ (note additional $n - 1$ index). The gradient elements of the transfer amplitude are then

$$\begin{aligned} \frac{\partial o}{\partial u_n} &= \frac{\partial}{\partial u_n} \langle \chi_{N_t} | \hat{\mathcal{U}}^{\text{ST}} | \psi_1 \rangle = \sum_{q=n-1}^{N_t-1} \left\langle \chi_{q+1} \left| \frac{\partial \hat{\mathcal{U}}_q^{\text{ST}}}{\partial u_n} \right| \psi_q \right\rangle \\ &= \alpha \sum_{q=n-1}^{N_t-1} \left\langle \chi_q \left| \frac{\partial \hat{H}_q^c}{\partial u_n} \right| \psi_q \right\rangle + \left\langle \chi_{q+1} \left| \frac{\partial \hat{H}_{q+1}^{c++}}{\partial u_n} \right| \psi_{q+1} \right\rangle, \end{aligned} \quad (5.17)$$

where we used that the control Hamiltonians are diagonal as in Sec. 4.4 so that

$$\begin{aligned} \frac{\partial \hat{\mathcal{U}}_q^{\text{ST}}}{\partial u_n} &= \frac{\hat{\mathcal{U}}_{q+1}^{c++/2}}{\partial u_n} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_q^{c/2} + \hat{\mathcal{U}}_{q+1}^{c++/2} \hat{\mathcal{U}}^d \frac{\partial \hat{\mathcal{U}}_q^{c/2}}{\partial u_n} \\ &= \alpha \left(\frac{\partial \hat{H}_q^{c++}}{\partial u_n} \hat{\mathcal{U}}_q^{\text{ST}} + \hat{\mathcal{U}}_q^{\text{ST}} \frac{\partial \hat{H}_q^c}{\partial u_n} \right). \end{aligned} \quad (5.18)$$

The control Hamiltonian derivatives can as in Eq. (5.13) be written as a recursion by the product rule now starting with base case $q = n - 1$. At recursion iteration $q > n - 1$, denoted in red, the quantities are

calculated in the following order (blue denotes quantities calculated in the previous recursion iteration),

$$\begin{aligned}
\frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{q-1} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} &= \frac{\partial \hat{\mathcal{U}}_{q-1}^{\text{ST}}}{\partial u_n} \psi_{q-1} + \hat{\mathcal{U}}_{q-1}^{\text{ST}} \frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{q-2} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} \\
&= \alpha \left(\frac{\partial \hat{H}_q^{c++}}{\partial u_n} \psi_q + \hat{\mathcal{U}}_{q-1}^{\text{ST}} \frac{\partial \hat{H}_{q-1}^c}{\partial u_n} \psi_{q-1} \right) + \hat{\mathcal{U}}_{q-1}^{\text{ST}} \frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{q-2} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} \\
&= \alpha \frac{\partial \hat{H}_q^{c++}}{\partial u_n} \psi_q + \hat{\mathcal{U}}_{q-1}^{\text{ST}} \left(\alpha \frac{\partial \hat{H}_{q-1}^c}{\partial u_n} \psi_{q-1} + \frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{q-2} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} \right), \tag{5.19a}
\end{aligned}$$

$$\frac{\partial \hat{H}_q^c}{\partial u_n} = \delta_{q,n} \frac{\partial V_n}{\partial u_n} + 2G_{1D} \Re \left(\psi_q^* \odot \frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{q-1} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} \right), \tag{5.19b}$$

$$\begin{aligned}
\frac{\partial \hat{H}_{q+1}^{c++}}{\partial u_n} &= \delta_{q+1,n} \frac{\partial V_n}{\partial u_n} + 2G_{1D} \Re \left((\psi_q^{++})^* \odot \hat{\mathcal{U}}^d \hat{\mathcal{U}}_q^{c/2} \right. \\
&\quad \left. \left\{ \alpha \frac{\partial \hat{H}_q^c}{\partial u_n} \psi_q + \frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{q-1} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} \right\} \right). \tag{5.19c}
\end{aligned}$$

In implementing these equations numerically, it is important to keep in mind that $\hat{\mathcal{U}}_q^{c/2} (\hat{\mathcal{U}}_{q-1}^{\text{ST}})$ contains the exponentiated state(s) $|\psi_q|^2$ ($|\psi_{q-1}|^2$ and $|\psi_{q-1}^{++}|^2$) rather than the states they are applied to.

Defining the “anti”-Kronecker delta $\delta_{i,j} = (1 - \delta_{i,j})$, the base case $q = n - 1$ values evaluate to

$$\frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{n-2} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} = 0, \quad \frac{\partial \hat{H}_{n-1}^c}{\partial u_n} = 0, \quad \frac{\partial \hat{H}_n^{c++}}{\partial u_n} = \delta_{n,1} \frac{\partial V_n}{\partial u_n}, \tag{5.20}$$

where the $\delta_{n,1}$ comes from the fact that there is no $\hat{\mathcal{U}}_0^{\text{ST}}$ operator. (similarly, $\partial \hat{H}_{N_t}^c / \partial u_n = 0$ because there is no $\hat{\mathcal{U}}_{N_t}^{\text{ST}}$ operator). Although already covered by Eqs. 5.19a-5.19c, the results for $q = n$ are calculated explicitly below since they are useful when implementing the gradient numerically,

$$\frac{\partial}{\partial u_n} \left(\prod_{j=n-1}^{n-1} \hat{\mathcal{U}}_j^{\text{ST}} \right) \psi_{n-1} = \alpha \left(\frac{\partial \hat{H}_n^{c++}}{\partial u_n} \psi_n + 0 \right) + 0 = \alpha \frac{\partial V_n}{\partial u_n} \psi_n \delta_{n,1}, \tag{5.21a}$$

$$\frac{\partial \hat{H}_n^c}{\partial u_n} = \frac{\partial V_n}{\partial u_n} + 2\delta_{n,1} G_{1D} |\psi_n|^2 \frac{\partial V_n}{\partial u_n} \Re(\alpha) = \frac{\partial V_n}{\partial u_n}, \tag{5.21b}$$

$$\frac{\partial \hat{H}_{n+1}^{c++}}{\partial u_n} = (2 + 2\delta_{n,1}) G_{1D} \Re \left((\psi_q^{++})^* \odot \hat{\mathcal{U}}^d \hat{\mathcal{U}}_q^{c/2} \left(\alpha \frac{\partial V_n}{\partial u_n} \psi_n \right) \right). \tag{5.21c}$$

The vanishing second term in Eq. (5.21b) is due to commutation properties of $\odot$,

$$\begin{aligned} \Re(\alpha \psi_n^* \odot \frac{\partial V_n}{\partial u_n} \odot \psi_n) &= \Re(\alpha \psi_n^* \odot \psi_n \odot \frac{\partial V_n}{\partial u_n}) = \Re(\alpha |\psi_n|^2 \frac{\partial V_n}{\partial u_n}) \\ &= |\psi_n|^2 \frac{\partial V_n}{\partial u_n} \Re(\alpha) = \frac{\delta t}{2} |\psi_n|^2 \frac{\partial V_n}{\partial u_n} \Re(i) = 0, \end{aligned} \quad (5.22)$$

and that both functions $|\psi|^2$ and $\partial V_n / \partial u_n$ are real such that $\Re(i) = 0$.

Numerical verification

To illustrate the validity of these somewhat cumbersome derivations, we now consider a simple “dummy” state transfer problem $|\psi_1\rangle = |\psi_{\text{tgt}}\rangle = |\phi_0; G_{1D}, u = 0\rangle$ in the potential $\hat{V}(x; u(t)) = \frac{1}{2}\omega^2(x - u(t))^2$. We take $\hbar = m = 1$, $\omega = 3$, $x \in [x_1, \dots, x_{128}] = [-5, \dots, 5]$, a constant control $u_n = 1$, and $T = 1$.

Figure 5.8 illustrates how the control Hamiltonians at indices $q > n$ are affected when varying the control u_n . Recall that this is due to the propagator nesting, e.g., $\hat{H}_n^c = V_n + |\hat{\mathcal{U}}_{1:n-1}^{\text{ST}} \psi_1|^2$. We observe that the derivative magnitudes grow with q . This reflects the “butterfly-effect” of modifying u_n — i.e., the steady departure from the “unperturbed” Hilbert space trajectory.

Figure 5.9 shows that the exact analytical gradients match their finite difference⁵ counterpart to (near) machine precision whereas the approximate analytical gradients has an error that scales increasingly with the product $\delta t \cdot G_{1D}$. In particular, at the experimentally realistic but relatively modest value $G_{1D}(700) = 2.5$ for the magnetic atom chip trap (c.f. Fig. 5.4), the approximate gradients are notably inaccurate even for the very granular $\delta t = 0.001$. For optical geometries with millions of atoms, $G_{1D} \gg 1$, these effects are therefore expected to be even more pronounced. Curiously, and currently remaining unexplained, there seems to be a slightly increased systematic error associated with the *inverse* of δt for the exact derivatives. This is likely due to a minor implementation error or similar. Further, it is pointed out that the $G_{1D} = 0$ results shown here are only “near-exact”. “Exact” because of the fortuitous coincidence in Sec. 4.5.3, but only “nearly” so because the backwards propagation of $|\chi_{N_i}\rangle$ is here numerically performed as in Ref. [264] (representative of the literature standard). In practice, of course, one does not use this more complicated stepping algorithm when $G_{1D} = 0$. Nevertheless, it is illustrative and legitimizing that this one change in numerics, despite being *formally* equivalent in the $G_{1D} = 0$ case, makes it so that the computed gradient is no longer fully aligned with the geometric one as was detailed to a somewhat lengthy extent in Sec. 4.6.

⁵ Finite differencing is itself associated with error $\epsilon_{\text{finite}} \sim \sqrt{\epsilon_{\text{mach}}} \sim 10^{-8}$ so a relative comparison to the analytical ones can only yield meaningful precision to order $\epsilon_{\text{mach}} / \epsilon_{\text{finite}} \sim 10^{-8}$

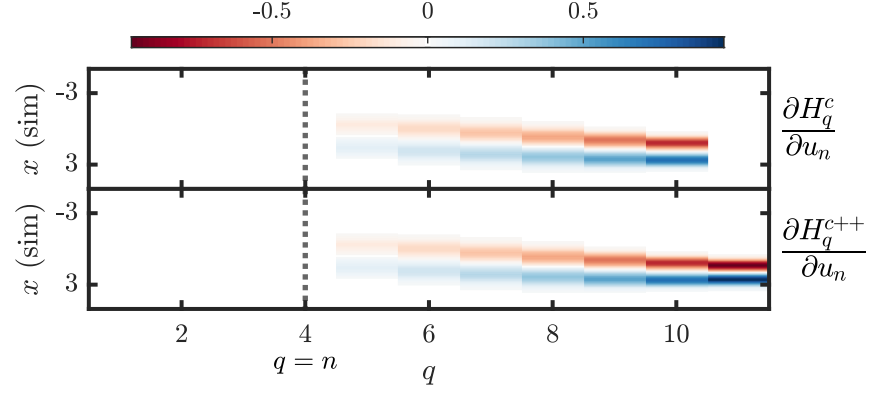


Figure 5.8: Example of calculating (5.19b)-(5.19c) for $n = 4$, $G_{1D} = 2.5$ and $\delta t = 0.1$. The first axis denotes the time index q and the color indicates the corresponding diagonal values of the control Hamiltonian derivatives (the off-diagonals are zero in the $\hat{x}$ -basis). Note that the large contributions at index $q = n$ (dotted line) due to $\partial V_n / \partial u_n$ have been subtracted for readability.

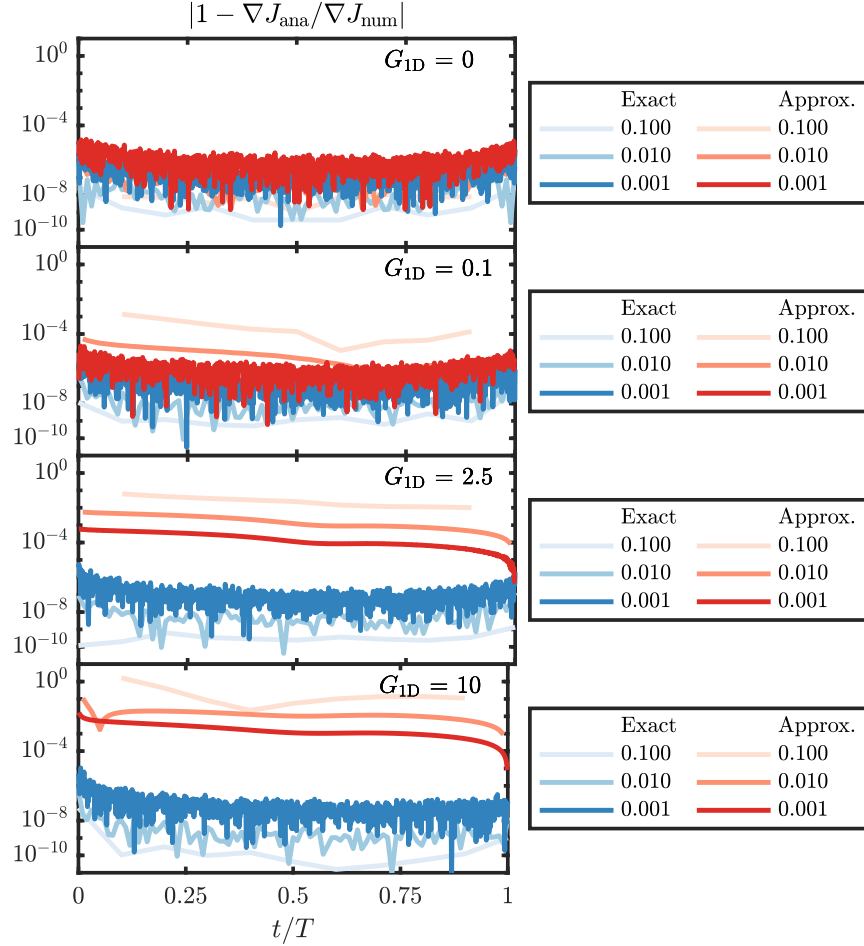


Figure 5.9: Analytical gradients (∇J_{ana}) relative to their numerical central finite difference approximation (∇J_{num}) with perturbation $\epsilon_{\text{pert}} = \epsilon_{\text{mach}}^{1/3}$. Exact (blue lines) and Approx. (red lines) refers to calculating the analytical gradient with Eqs. (5.17)–(5.19c) or Eqs. (4.38), respectively, for different values of δt as shown in the legend entries. The approximate analytical gradients were calculated in the QENGINE [1] as also detailed in Ref. [264] whereas the exact analytical gradients were calculated in a separate MATLAB implementation.

Finally, with results similar to those in Secs. 4.4–4.5 where $G_{1D} = 0$, this verifies the derivations presented here. The consequences there also follow here. There is, however, an additional set of observations to be made in this situation. As opposed to the fortuitous circumstances detailed in Sec. 4.5.3 when $G_{1D} = 0$, the optimize-then-discretize and discretize-then-optimize gradients never exactly coincide when $G_{1D} \neq 0$. To my knowledge, these generalized discretize-then-optimize gradients are not present in the literature and the optimize-then-discretize approach is always followed. The implication is that all similarly natured derivative-based control applications [1, 5, 75, 161, 162, 165, 167, 197, 264, 297, 298] regarding meanfield BEC have, to date, been optimizing with approximate gradients cf. Fig. 5.9 and their associated hampered performance cf. Figs. 4.6–4.7. These results could therefore potentially be a useful addition to the theoretical tool box in solving such problems and addressing, e.g., the questions in Sec. 5.2.

This concludes our discussion on analytically exact derivatives and quantum optimal control techniques as a whole. We now turn to more application focused works in the following chapters.

TWO-PARTICLE $\sqrt{\text{SWAP}}$ COLLISION GATES

AMONG other requirements listed in Sec. 2.1.1, quantum computing and processing tasks require a set of universal gates that can be implemented ideally with perfect fidelity and preferably near the fundamentally allowed minimal duration. The universal set usually consists in a number of single-qubit gates and an entangling two-qubit gate. Typically, the single-qubit gates can be routinely implemented consistent with error correction requirements whereas realizing the same for two-qubit gates is more challenging. Its operation is therefore an obvious candidate for quantum optimal control.

6.1 INTRODUCTION TO THE MANUSCRIPT

In the work of Ref. [2] we explore and numerically implement the two-qubit $\sqrt{\text{SWAP}}$ entangling gate, i.e. the unitary $\hat{U}_{\sqrt{\text{SWAP}}}$, based on collisional (short-range contact) interactions in ultracold neutral atoms using quantum optimal control techniques. Quite generally, collision gates are effected by merging two indistinguishable atoms into the same spatial (collision) region, waiting for appropriate entanglement to be generated by the interactions, and finally separating the atoms anew. This procedure is shown schematically in Fig. 6.1 for the $\sqrt{\text{SWAP}}$ gate.

The work encompasses several analytical and numerical details at least one of which are unaccounted for in similar previous works [72, 109, 140, 145, 171, 299]. Before giving a more quantitative overview we briefly summarize some of these considerations and results as follows. We thoroughly analyze the system, interaction effects, and gate in an independent-particle picture. Establishing a suitable notation, these analyses are then naturally extended and connected to the case of genuine two-particle wave functions which are much harder to visualize and intuit. In the numerics we include both singlet and triplet state components, the relative phases of which is what in fact drives the gate. For optimal gate implementation, we show for the usual stage-wise protocol (merge-wait-separate) that fidelity is not a good figure of merit. Additionally, it is shown that the stage-wise protocol is both more imprecise and slower in terms of $T_{\min}^{F=0.99}$ than optimizing the full gate directly and this can be interpreted in terms of restrictively conditioned Hilbert space trajectories.

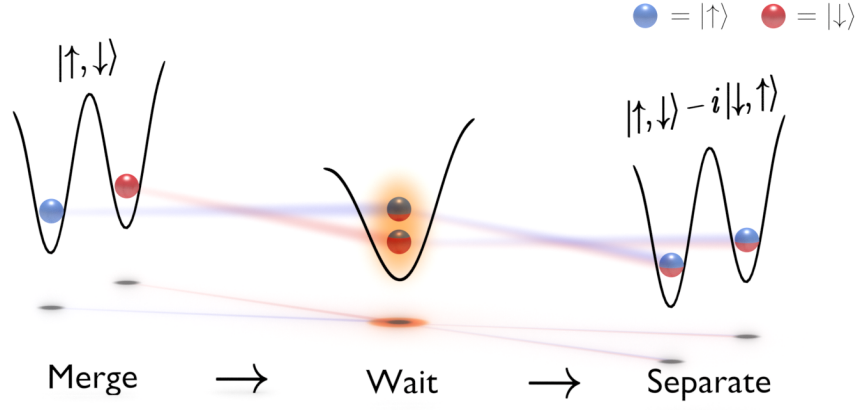


Figure 6.1: From Ref. [2]. Schematic illustration of the typical merge-wait-separate sequence implementing the $\sqrt{\text{SWAP}}$ gate. Completing the sequence transfers a state of initially opposite spins into a spin-entangled state. Initially aligned spins are instead imparted a phase $e^{-i\pi/4}$. These basis state mappings collectively implement the unitary $\mathcal{U}_{\sqrt{\text{SWAP}}}$ gate.

6.2 MANUSCRIPT RESULTS

The time evolution is governed by the two-particle Hamiltonian Eq. (2.59)

$$\hat{H}^{(2)} = \sum_{i=1}^2 \hat{h}(x_i) + g_{1D} \delta(x_1 - x_2), \quad (6.1)$$

where $\hat{h}(x_i) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_i^2} + V(x_i)$ is the single-particle Hamiltonian for particle $i = 1, 2$, V is the controllable trapping potential, and g_{1D} the collision interaction strength that generates the entanglement.

In this work we consider the bosonic rubidium 87 atomic species with general, properly symmetrized states on the form¹

$$|\Phi^+\rangle = |\Psi^+\rangle |\chi^+\rangle, \quad |\Phi^{\downarrow\downarrow}\rangle = |\Psi^+\rangle |\chi^{\downarrow\downarrow}\rangle, \quad |\Phi^{\uparrow\uparrow}\rangle = |\Psi^+\rangle |\chi^{\uparrow\uparrow}\rangle, \quad (6.2)$$

$$|\Phi^-\rangle = |\Psi^-\rangle |\chi^-\rangle, \quad (6.3)$$

where $|\Psi^\pm\rangle$ are symmetric (+) and anti-symmetric (−) spatial wave functions, $|\chi^\pm\rangle = |\uparrow\rangle_1 |\downarrow\rangle_2 \pm |\downarrow\rangle_1 |\uparrow\rangle_2$ are spin states with the same identification, and $|\chi^{\uparrow\uparrow}\rangle = |\uparrow\rangle_1 |\uparrow\rangle_2$ and $|\chi^{\downarrow\downarrow}\rangle = |\downarrow\rangle_1 |\downarrow\rangle_2$ — subscripts denote particle labels. Thus, Ψ^+ is uniquely associated with the spin triplet states and we shall also therefore refer to it as a triplet state and similar for Ψ^- .

The spin and spatial degrees of freedom separate exactly in model Eq. 6.1 cf. Sec. 2.4 and we can therefore disregard explicit inclusion

¹ Proper normalization is implied throughout. Additionally, the analyses and results presented here are valid also for fermions upon a trivial interchange of symmetry requirements: $|\Phi^{\uparrow\uparrow}\rangle = |\Psi^-\rangle |\chi^{\uparrow\uparrow}\rangle$, $|\Phi^{\downarrow\downarrow}\rangle = |\Psi^-\rangle |\chi^{\downarrow\downarrow}\rangle$, and $|\Phi^\pm\rangle = |\Psi^\pm\rangle |\chi^\mp\rangle$.

of the spin in numerical implementation. This is despite the fact that these are, in fact, the qubit states to be entangled — the entanglement generated for a collision-pair in the auxiliary real space carries over to the qubits due to exchange symmetry requirements. Nevertheless, although the qubits are thus just “along for the ride”, they must be included in some of the analytical analyses as follows.

6.2.1 Gate description

It pays to establish a subscript notation relating to the *independent-particle picture* where the spatial state remains on the simple form

$$|\Psi_{a,b}^{\pm}\rangle = |a\rangle_1 |b\rangle_2 \pm |b\rangle_1 |a\rangle_2, \quad (6.4)$$

where $a(x,t) = \langle x|a(t)\rangle$ and $b(x,t) = \langle x|b(t)\rangle$ are generic single particle states. By now adding the physically allowed symmetrized states in various ways we construct a convenient computational basis that uniquely designates the single particle states with a definite spin regardless of particle label,

$$|\uparrow_a, \downarrow_b\rangle \equiv |\uparrow_a\rangle_1 |\downarrow_b\rangle_2 + |\downarrow_b\rangle_1 |\uparrow_a\rangle_2 = |\Phi_{a,b}^+\rangle + |\Phi_{a,b}^-\rangle, \quad (6.5a)$$

$$|\downarrow_a, \uparrow_b\rangle \equiv |\downarrow_a\rangle_1 |\uparrow_b\rangle_2 + |\uparrow_b\rangle_1 |\downarrow_a\rangle_2 = |\Phi_{a,b}^+\rangle - |\Phi_{a,b}^-\rangle, \quad (6.5b)$$

$$|\uparrow_a, \uparrow_b\rangle \equiv |\uparrow_a\rangle_1 |\uparrow_b\rangle_2 + |\uparrow_b\rangle_1 |\uparrow_a\rangle_2 = |\Phi_{a,b}^{\uparrow\uparrow}\rangle = |\Psi_{a,b}^+\rangle |\chi^{\uparrow\uparrow}\rangle, \quad (6.5c)$$

$$|\downarrow_a, \downarrow_b\rangle \equiv |\downarrow_a\rangle_1 |\downarrow_b\rangle_2 + |\downarrow_b\rangle_1 |\downarrow_a\rangle_2 = |\Phi_{a,b}^{\downarrow\downarrow}\rangle = |\Psi_{a,b}^+\rangle |\chi^{\downarrow\downarrow}\rangle. \quad (6.5d)$$

The $\sqrt{\text{SWAP}}$ operation thus amounts to the following basis state mappings,

$$|\uparrow_a, \downarrow_b\rangle \rightarrow \hat{U}_{\sqrt{\text{SWAP}}} |\uparrow_a, \downarrow_b\rangle = |\uparrow_a, \downarrow_b\rangle - i |\downarrow_a, \uparrow_b\rangle, \quad (6.6a)$$

$$|\downarrow_a, \uparrow_b\rangle \rightarrow \hat{U}_{\sqrt{\text{SWAP}}} |\downarrow_a, \uparrow_b\rangle = -|\uparrow_a, \downarrow_b\rangle - i |\downarrow_a, \uparrow_b\rangle, \quad (6.6b)$$

$$|\uparrow_a, \uparrow_b\rangle \rightarrow \hat{U}_{\sqrt{\text{SWAP}}} |\uparrow_a, \uparrow_b\rangle = e^{-i\pi/4} |\uparrow_a, \uparrow_b\rangle, \quad (6.6c)$$

$$|\downarrow_a, \downarrow_b\rangle \rightarrow \hat{U}_{\sqrt{\text{SWAP}}} |\downarrow_a, \downarrow_b\rangle = e^{-i\pi/4} |\downarrow_a, \downarrow_b\rangle, \quad (6.6d)$$

which entangles initially anti-aligned spins and imparts a phase $e^{-i\pi/4}$ on aligned spins². The goal is thus to effect this gate by using optimal control. Under reasonable assumptions (isotropic scattering lengths), a protocol implementing just $|\uparrow_a, \downarrow_b\rangle \rightarrow \hat{U}_{\sqrt{\text{SWAP}}} |\uparrow_a, \downarrow_b\rangle$ also correctly maps all the remaining basis states. We therefore restrict our attention to this state transfer problem which can for our purposes be conveniently written in terms of singlet and triplet states,

$$|\psi_{\text{ini}}\rangle = |\Phi_{a,b}^+\rangle + |\Phi_{a,b}^-\rangle \rightarrow |\psi_{\text{tgt}}\rangle = |\Phi_{a,b}^+\rangle + e^{i\pi/2} |\Phi_{a,b}^-\rangle. \quad (6.7)$$

² Usually the $\sqrt{\text{SWAP}}$ imparts a phase $e^{i\phi} = 1$ on aligned spins but this slightly different definition is inconsequential since entanglement generation is the express purpose here.

With these identifications it becomes evident that relative phases between singlet and triplet components is what determines the spin/space correlations and thus by corollary the entanglement. In previous theoretical optimal control efforts [72, 299], however, only the triplet component was considered and this means they are not entirely implementing the correct entangling dynamics (it can be considered an approximation, see the manuscript).

For these mappings to be possible, there clearly has to be some way to alter the relative phase $1 \rightarrow e^{i\pi/2}$ between the singlet and triplet components. This is realized as follows. The properly symmetrized single particle states have readily calculable energies

$$\langle \Psi_{a,b}^- | \hat{H} | \Psi_{a,b}^- \rangle = E_a + E_b, \quad (6.8)$$

$$\langle \Psi_{a,b}^+ | \hat{H} | \Psi_{a,b}^+ \rangle = E_a + E_b + U_{a,b}, \quad (6.9)$$

with E_a and E_b being the usual single particle energies and the interaction energy³

$$U_{a,b} \equiv 2 \int_{-\infty}^{\infty} |a(x)|^2 |b(x)|^2 g_{1D}(x) dx. \quad (6.10)$$

A central point is that *all* singlet states remain unaffected by the interaction since $\Psi^-(x_1, x_2 = x_1) = 0$ vanishes due to anti-symmetry which, incidentally, are the only coordinates where $g_{1D}\delta(x_1 - x_2)$ is non-zero. The singlet and triplet states are thus always subject to a differential energy, here $\Delta E_{a,b}^\pm = U_{a,b}$, and this is what enables control over their relative phases.

Concretely, assuming that $|a(t)\rangle$ and $|b\rangle$ are adiabatic eigenstates of $\hat{h}$, the approximate time evolution of a two-particle state prepared as $|\uparrow_a, \downarrow_b\rangle$ is easily described in the singlet/triplet decomposition and subsequently converted back to the computational basis where a global phase is eliminated,

$$|\Phi_{a,b}(t)\rangle \approx |\Phi_{a,b}^+\rangle + e^{i\alpha} |\Phi_{a,b}^-\rangle \quad (6.11a)$$

$$\rightarrow \cos\left(\frac{\alpha}{2}\right) |\uparrow_a, \downarrow_b\rangle - i \sin\left(\frac{\alpha}{2}\right) |\downarrow_a, \uparrow_b\rangle. \quad (6.11b)$$

The relative phase advancement in this two-component Rabi dynamics is illustrated in Fig. 6.2. Crucially, α is advanced by the interactions

$$\alpha(t) = \hbar^{-1} \int_0^t U_{a(t'), b(t')} dt', \quad (6.11c)$$

and the rate of entanglement generation, $\dot{\alpha} \propto U_{a(t), b(t)} \geq 0$, is determined by the spatial overlap and collisional coupling. This brings with it a coincidental set of central observations, namely

³ Note that we already calculated this in Eq. (2.41) using second quantization. Here, we have to keep $g_{1D}(x)$ inside the integral because it is a function of x , see Sec. 2.4 and the manuscript.

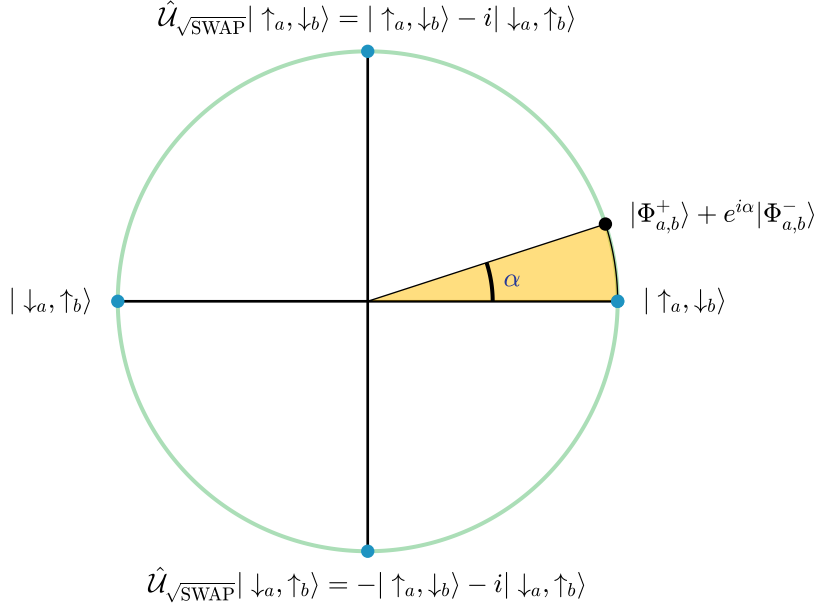


Figure 6.2: Adapted from Ref. [2]. Phase plane. The relative phase α between singlet-triplet components determines the spin-space correlations.

1. a and b must be in contact to engage the phase evolution, $\dot{\alpha} > 0$,
2. the phase evolution can march only forward, $\dot{\alpha} \geq 0$, and
3. the initial and final states must be interaction-free such that no further entanglement is generated, $\dot{\alpha} = 0$.

The last observation is due to the intended use in quantum computing. That is, it is not enough to simply enter $\hat{U}_{\sqrt{\text{SWAP}}}|\uparrow_a, \downarrow_b\rangle$, correspondingly $\alpha = \pi/2$, for any a and b because a non-vanishing density overlap will advance $\alpha > \pi/2$ “irreparably” since there is no way to revert to smaller values $\dot{\alpha} < 0$ other than waiting for full revolutions in the phase plane. Collectively, this means that the atoms must be merged from an interaction-free configuration into an intermediate interaction-engaged one and then separated back into the original, all the while making sure to exactly procure $\alpha(T^f) = \hbar^{-1} \int_0^{T^f} U_{a(t'), b(t')} dt' = \pi/2$ at the final duration T^f . This must therefore be taken into consideration when formulating the control problem of Eq. (6.7) in any trapping geometry. Since the aforementioned previous theoretical works [72, 299] did not include the singlet component, this otherwise central detail is absent there.

Genuine Two-Particle Eigenstates

Lastly, we must reconcile that the analyses above were predicated on an independent-particle picture where $|\Psi_{a,b}^\pm\rangle = |a\rangle_1 |b\rangle_2 \pm |b\rangle_1 |a\rangle_2$ follow eigenstates of $\hat{h}$. We can, however, keep a meaningful and rigorous

reference to these simple single-particle states by the following line of reasoning⁴.

Recall that the calculations in Eq. (2.42) expressly showed that individual, symmetrized single-particle energy eigenstates $\epsilon_i = \langle x|E_i\rangle$ are *not* also generally eigenstates for the interaction. This could for example be seen from the instantaneous time evolution of the corresponding multi-particle state components, here on the form $|1_{E_n}, 1_{E_m}\rangle$, which experience an instantaneous occupational mixing proportional to

$$\sum_{ijkl} U_{ijkl} |n_{E_i} + 1, n_{E_j} + 1, n_{E_k} - 1, n_{E_l} - 1\rangle \dots, \quad (6.12)$$

where $U_{ijkl} = \int \epsilon_i^* \epsilon_j^* \epsilon_k \epsilon_l g_{1D}(x) d^d r$. In the first quantized picture, this entails that the genuine two-particle spectrum $\hat{H}^{(2)} |\tilde{\Psi}_i\rangle = \tilde{E}_i |\tilde{\Psi}_i\rangle$ (note $\sim$ notation) is, *ab initio*, a non-trivial mixture $|\tilde{\Psi}_i\rangle = \sum_{k,l,s} c_{k,l}^s |\Psi_{E_k,E_l}^s\rangle$ where $k, l = 0, 1, \dots$ are single-particle eigenenergy indices and $s \in \{+, -\}$ the symmetry. Because $|\tilde{\Psi}_i\rangle$ must obey overall symmetry it follows that either all $c_{k,l}^+ = 0$ or all $c_{k,l}^- = 0$ and the expansion reduces to one of the possible forms, $s = s'$, such that $|\tilde{\Psi}_i^{s'}\rangle = \sum_{k,l} c_{k,l}^{s'} |\Psi_{E_k,E_l}^{s'}\rangle$. Suppose momentarily that the interactions are sufficiently weak, then the expansion for any given i will be dominated by a single coefficient $|c_{k',l'}^{s'}| \approx 1$ for some $k = k', l = l'$. We then associate for each i the unique identification $i \leftrightarrow (k', l', s')$ and may conveniently write

$$|\tilde{\Psi}_{k,l}^{s'}\rangle = c_{k',l'}^{s'} |\Psi_{k',l'}^{s'}\rangle + \delta |\tilde{\Psi}^{s'}\rangle, \quad (6.13)$$

where $\delta |\tilde{\Psi}^{s'}\rangle$ contains the terms of same symmetry with indices $k \neq k', l \neq l'$. For the symmetric (triplet) wave functions, the “mixing term”/perturbation takes values $|\delta \tilde{\Psi}^+| \geq 0$ that are small while $\delta \tilde{\Psi}^- = 0$ vanishes identically for the anti-symmetric (singlet) wave functions as they are unaffected by the interactions. This means that for any fixed pair of k, l (now dropping the prime), the genuine two-particle eigenstates will satisfy $\hat{H}^{(2)} |\tilde{\Psi}_{k,l}^\pm\rangle = \tilde{E}_{k,l}^\pm |\tilde{\Psi}_{k,l}^\pm\rangle$ with

$$\tilde{U}_{k,l} = \Delta \tilde{E}_{k,l}^\pm = \tilde{E}_{k,l}^+ - \tilde{E}_{k,l}^- \geq 0. \quad (6.14)$$

In preparing the genuine two-particle state as e.g. $|\tilde{\Phi}\rangle = |\tilde{\Phi}_{k,l}^+\rangle + |\tilde{\Phi}_{k,l}^-\rangle$, this differential energy between singlet and triplet components facilitates the relative phase evolution by the same mechanisms as before — $|\tilde{\Phi}\rangle$ remains on this two-component with an evolution of the relative phase $\dot{\alpha} \propto \tilde{U}_{k,l}$. Thus, the gate analysis given above in terms of the independent-particle picture becomes exact when replacing all quantities by their $\sim$ version. Inspired by this, we may then generalize these insights to arbitrary interaction strengths.

⁴ The following paragraph can be skipped without much loss of continuity. The final results are displayed in Eqs. 6.15.

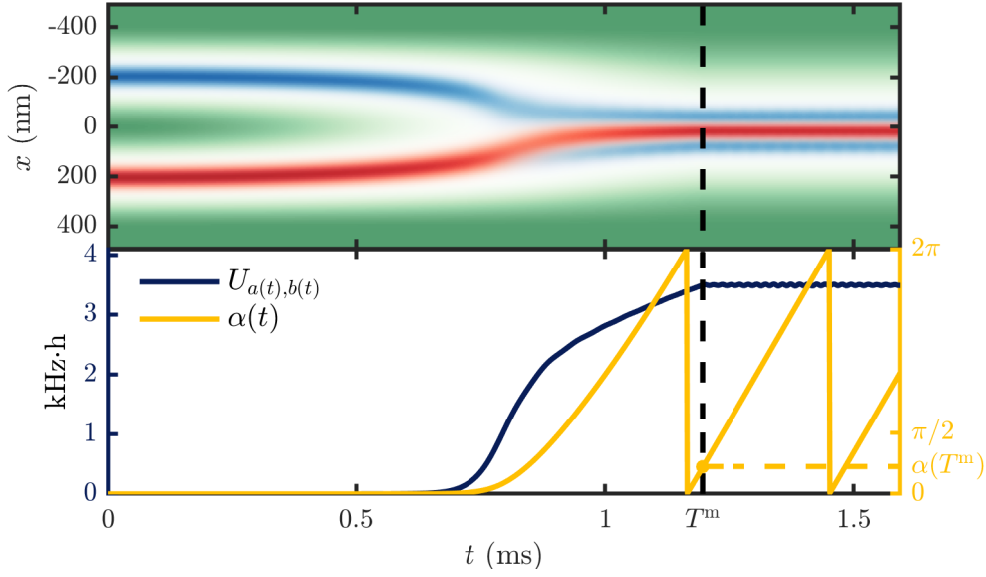


Figure 6.3: Adapted from Ref. [2]. Adiabatic merging $|a(0)\rangle|b(0)\rangle = |Lg\rangle|Rg\rangle \rightarrow |e\rangle|g\rangle = |a(T^m)\rangle|b(T^m)\rangle$. Top: The instantaneous densities $|a(x,t)|^2$ (blue) and $|b(x,t)|^2$ (red) remain eigenstates and are merged by a linear adiabatic transformation of the potential from a double- to single-well configuration (green denote high potential values). Bottom: The interaction engages when the states begin to overlap and the phase evolves counter-clockwise as $\dot{\alpha} \sim U_{a(t),b(t)}$ in the phase plane of Fig. 6.2. After the merge, $t > T^m$, the trap is held static and the rate of change is constant. [Note density colormaps visually fades and do not reflect the full extent of the states].

We define the genuine two-particle eigenstate $|\tilde{\Psi}_{k,l}^+\rangle$ as the one that turns into the corresponding symmetrized single-particle eigenstate

$$|\tilde{\Psi}_{k,l}^-\rangle = |\Psi_{k,l}^-\rangle, \quad (6.15a)$$

$$|\tilde{\Psi}_{k,l}^+\rangle \rightarrow |\Psi_{k,l}^+\rangle, \quad (6.15b)$$

in the limit ($\rightarrow$) of vanishing interactions. Vanishing interactions are either due to vanishing spatial overlap, no coupling $g_{1D} = 0$, or spatial anti-symmetry requirements. Thus, we retain reference to the simple single-particle states. The actual states $|\tilde{\Psi}_{k,l}^\pm\rangle$ are found by numerical diagonalization of $\hat{H}^{(2)}$, yet the analytical analyses were important in understanding and formulating the gate description.

It is finally noted that in applying optimal control we do not intend to impose adiabatic following for all t and, in general, many instantaneous eigenstates $|\tilde{\Psi}_{k,l}^\pm\rangle$ will be populated. This fact, and the overall detailed connection to the genuine two-particle states, are not present in previous works of primarily experimental nature [109, 140, 145, 171].

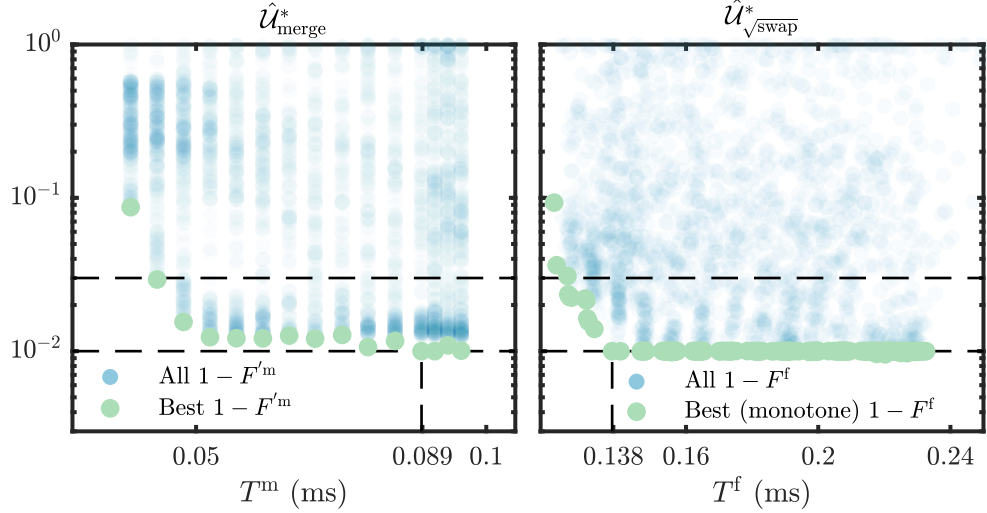


Figure 6.4: Adapted from Ref. [2]. Optimizing merging in explicit stage-wise protocol form (left) or the full gate (right). We find respective minimal durations $T_{\min}^{m, F'=0.99} = 0.089$ ms and $T_{\min}^{f, F=0.99} = 0.138$ ms.

6.2.2 Example geometry

We consider the optical lattice in Eq. (2.8)

$$V = V_{\text{merge}}(x; V_0(t), \beta(t), \theta(t)), \quad (6.16)$$

with three control parameters influencing the potential as an example geometry since this was used in the first experimental realizations [140, 171]. In the independent-particle picture, the transfer performed there was achieved by adiabatically merging a pair of lattice sites into a single site, implementing the mapping $|Lg\rangle |Rg\rangle \rightarrow |e\rangle |g\rangle$. The notation Lg (Rg) refers to the ground state in the left (right) well and e (g) to the first excited (ground) state in the merged well where the interaction primarily takes place. This is illustrated numerically in Fig. 6.3. Here, the function triplet $\{V_0(t), \beta(t), \theta(t)\}$ is found by optimal control which significantly speeds up the transfer times and fidelities. The unentangled initial state and entangled target state, both interaction-free, for the full gate are

$$|\psi_{\text{ini}}\rangle = |\uparrow_{Lg}, \downarrow_{Rg}\rangle \rightarrow |\psi_{\text{tgt}}^f\rangle = \hat{U}_{\sqrt{\text{SWAP}}} |\uparrow_{Lg}, \downarrow_{Rg}\rangle. \quad (6.17)$$

The numerically attained genuine two-particle states forming these are shown in the manuscript appendix.

Stage-wise protocol

The typically proposed control protocol [72, 109, 140, 145, 171, 299] for the gate implementation in Eq. (6.17) is on an explicitly stage-wise form, $\hat{U}_{\sqrt{\text{SWAP}}} = (\hat{U}_{\text{separate}} \cdot \hat{U}_{\text{wait}}) \cdot \hat{U}_{\text{merge}}^*$. Assuming an appropriate

high-fidelity merging stage $\hat{\mathcal{U}}_{\text{merge}}^*$ (adiabatic or optimized with optimal control) into an intermediate interaction-engaged state, here on the form $|\psi_{\text{tgt}}^{\text{m}}\rangle = |\Phi_{e,g}^+\rangle + e^{i\alpha} |\Phi_{e,g}^-\rangle$, the two final stages ($\hat{\mathcal{U}}_{\text{separate}} \cdot \hat{\mathcal{U}}_{\text{wait}}$) can be tailored individually to complete the gate.

To see how this can be done, it is helpful to record the explicit stage-wise evolution thusly

$$|\psi(T^{\text{f}})\rangle = (\hat{\mathcal{U}}_{\text{separate}} \cdot \hat{\mathcal{U}}_{\text{wait}}) |\psi(T^{\text{m}})\rangle = (\hat{\mathcal{U}}_{\text{separate}} \cdot \hat{\mathcal{U}}_{\text{wait}}) \cdot \hat{\mathcal{U}}_{\text{merge}}^* |\psi_{\text{ini}}\rangle, \quad (6.18)$$

where T^{f} and T^{m} are the full-gate and merging stage durations, respectively. As indicated in Fig. 6.3, the acquired phase after the merge, $\alpha(T^{\text{m}})$, is generally variable. In the waiting stage $\hat{\mathcal{U}}_{\text{wait}}$ we hold the trap static for some duration T^{w} to be determined shortly. During this stage, the two-component form of $|\psi_{\text{tgt}}^{\text{m}}\rangle$ undergoes Rabi dynamics and accumulates phase at a constant rate. It is important to reiterate, however, that we are not interested in waiting to fully enter the $\sqrt{\text{SWAP}}$ 'ed interaction-engaged state, $\hat{\mathcal{U}}_{\sqrt{\text{SWAP}}} |\uparrow_e, \downarrow_g\rangle$, since phase is also accumulated in separating the atoms which would finally lead to $|\psi(T^{\text{f}})\rangle \neq \hat{\mathcal{U}}_{\sqrt{\text{SWAP}}} |\uparrow_{Lg}, \downarrow_{Rg}\rangle$. We must therefore, instead, “prematurely” return the atoms to their initial positions such that exactly the remaining phase is picked up during the separation stage. This is most easily done by letting $\hat{\mathcal{U}}_{\text{separate}}$ be constructed by running the $\hat{\mathcal{U}}_{\text{merge}}^*$ control in reverse — this process then naturally has duration $T^{\text{s}} = T^{\text{m}}$ and picks up an identical phase $\alpha(T^{\text{m}})$. Finally, this conditions the T^{w} duration to fulfill $\alpha(T^{\text{f}}) = 2\alpha(T^{\text{m}}) + \alpha(T^{\text{w}}) = \pi/2$. In this way, it is possible to implement the full gate as schematized in Fig. 6.1 based solely on $\hat{\mathcal{U}}_{\text{merge}}^*$ with a duration of $T^{\text{f}} = T^{\text{m}} + T^{\text{w}} + T^{\text{s}} = 2T^{\text{m}} + T^{\text{w}}$.

Full gate protocol

Recall, however, that optimal control does not explicitly require this type of *a priori* specification of the physical mechanisms and processes (i.e. here stage-wise decomposition) cf. Sec. 3.8. In instances where we do wish to impose such restrictions, this should rather be done through, e.g., additional cost functional inclusions. Here, however, the explicit stage-wise decomposition is not a fundamental problem property and it is therefore much more attractive to optimize the full-gate implementation directly, i.e. realizing $|\psi_{\text{tgt}}^{\text{f}}\rangle = |\psi(T^{\text{f}})\rangle = \hat{\mathcal{U}}_{\sqrt{\text{SWAP}}}^* |\psi_{\text{ini}}\rangle$ as originally stated in Eq. (6.17). To our knowledge, this had not been done before.

Optimization results

We optimize the stage-wise and full gate protocols separately. For the latter, we use the merging-optimized stage-wise protocols as seeds and find much better results as shown in Fig. 6.4. The seed that optimized to the $T_{\text{min}}^{f,F=0.99} = 0.138 \text{ ms}$ solution had initially only $F^{\text{m}} = 0.93$ at

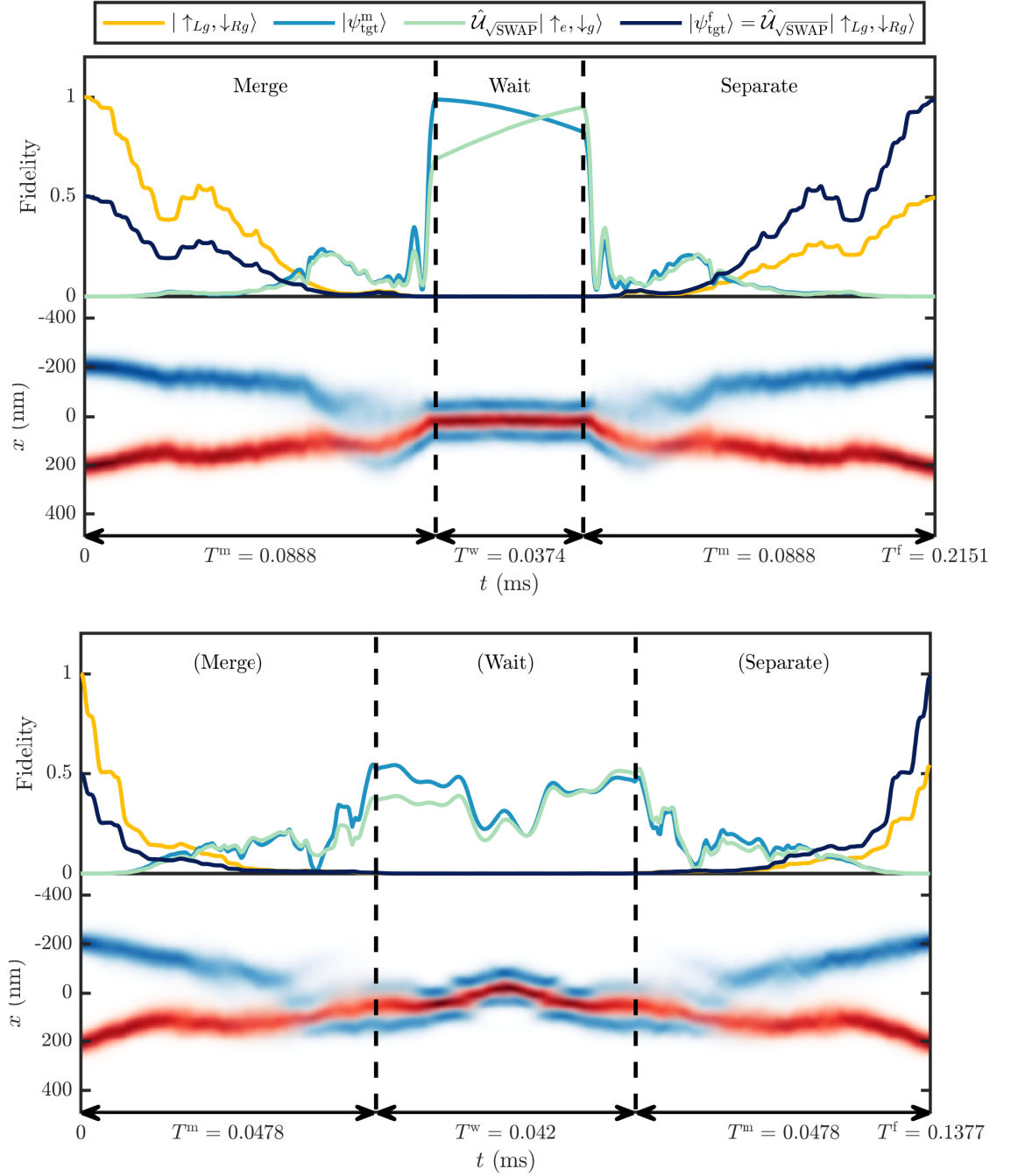


Figure 6.5: Adapted from Ref. [2]. Implementing the $\sqrt{\text{SWAP}}$ gate with optimal control, showing dynamical evolutions for the genuine two-particle fidelities and the single-particle densities (independent-particle picture) at $T_{\min}^{m, F'=0.99}$ and $T_{\min}^{f, F=0.99}$. Top panels: Explicit stage-wise control protocol $\hat{U}_{\sqrt{\text{SWAP}}} = (\hat{U}_{\text{separate}} \cdot \hat{U}_{\text{wait}}) \cdot \hat{U}_{\text{merge}}^*$. The optimized merging stage yields the intermediate target fidelity $F'(T^m) = 0.99$, but the final fidelity is $F(T^f) = 0.983$. Bottom panels: Less restrictive full-gate optimized protocol $\hat{U}_{\sqrt{\text{SWAP}}}^*$ yielding a final fidelity $F(T^f) = 0.99$. Here, the parenthetical stages and denoted times refer only to the numbers for the seed (for comparative purposes) and not the optimized results.

$T^m = 0.0478$ ms, showing that the stage-wise protocol is unnecessarily restrictive. Note that, in optimizing the merging stage, the fidelity is not the best figure of merit because any $|\Phi_{e,g}^+\rangle + e^{i\alpha} |\Phi_{e,g}^-\rangle$ is acceptable so long as $\alpha \leq \pi/4$ (if it exceeds, then we must wait for a full, slow revolution in the phase plane). A better objective would be the total population in the two-component singlet-triplet form, $F'^m = |\langle \Phi_{e,g}^+ | \Phi(T^m) \rangle|^2 + |\langle \Phi_{e,g}^- | \Phi(T^m) \rangle|^2$. Here we simply optimize the fidelity with a *target phase*, $\alpha_{\text{tgt}} \approx 0.33$, while reporting and using F'^m as a termination condition in place of fidelity.

Figure 6.5 shows examples of optimized $(\hat{\mathcal{U}}_{\text{separate}} \cdot \hat{\mathcal{U}}_{\text{wait}}) \cdot \hat{\mathcal{U}}_{\text{merge}}^*$ and $\hat{\mathcal{U}}_{\sqrt{\text{SWAP}}}^*$ dynamics. For the stage-wise protocol, we precisely enter a two-component singlet-triplet state exhibiting Rabi-dynamics in the interval $T^m \leq t \leq T^m + T^w$. However, even though the merging population reaches $F'(T^m) = 0.99$, the final fidelity $F(T^f) = 0.983$ deteriorates slightly because errors are enhanced in propagating the waiting and separation stages. For the full-gate protocol, we do not require exclusive population in just the two lowest single-particle eigenstates (in the independent-particle picture) and the particles “slosh” around while remaining largely overlapped. The genuine two-particle state is therefore never on a simple two-component singlet-triplet form at intermediate times which is evident from the absence of extended Rabi-dynamics for all $0 < t < T$. This allows a final fidelity $F(T^f) = 0.99$ as well as much faster and less restrictive overall operation — the particles are, e.g., not required to be intermittently (de-)accelerated (to) from rest which can be difficult to achieve. These and similar effects are responsible for the sub-optimal plateau in Fig. 6.4 for $\hat{\mathcal{U}}_{\sqrt{\text{SWAP}}}^*$.

6.3 PERSPECTIVES

We have demonstrated — taking into account many details — that optimal control techniques can implement a fundamental collisional two-qubit gate with high fidelity for neutral atom quantum computing [129]. These studies could be useful in relation to contemporary developments where it has become possible to trap and exploit alkaline-earth-like species (AELS) such as ytterbium and strontium [116–118, 300] rather than just alkalis like rubidium. AELS store the logical qubit information in the physical nuclear spin states and the electronic degrees of freedom are utilized as auxiliary means for state-selective transport and readout in similar architectures [127, 132, 301]. State-selective internal and external interactions allow potentially faster gate times and enhanced control. In particular, certain (e.g. ytterbium) isotopes possess orders of magnitude larger collisional interaction couplings (proportional to increased scattering lengths). Along similar lines, one could envision speeding up the gate further by maximizing scattering lengths by Feshbach resonance tuning [302].

There are, however, two crucial weaknesses to collision-based gates over the currently much more promising and actively researched Rydberg-based gates [76, 119, 303–309]. Briefly, the target qubit mapping in e.g. a CPHASE Rydberg gate is conditioned on whether or not the control qubit is momentarily driven into a specified Rydberg state (i.e. an internal state with very high principal quantum number, $n \sim \mathcal{O}(10^1) - \mathcal{O}(10^2)$) and thus blocking excitation due to the transition being shifted out of resonance.

The first weakness is that collisional interactions are relatively weak compared to the dipole-dipole Rydberg interactions, thus yielding much longer entangling gate times (roughly on order ms as found here compared to μs [143, 310, 311]). The second is that external motional degrees of freedom are much harder to reliably manipulate than internal states. This is true both in experiment and numerical simulations for a variety of reasons. For example, let us consider applying the full $\sqrt{\text{SWAP}}$ operation to an arbitrary qubit pair in the register

$$\hat{\mathcal{U}}_{\sqrt{\text{SWAP}}}^{(i,j)} |q_1, \dots, q_i, q_j, \dots, q_N\rangle = |q_1, \dots, \hat{\mathcal{U}}_{\sqrt{\text{SWAP}}}^{(i,j)}(q_i, q_j), \dots, q_N\rangle,$$

with the architecture proposed in Ref. [129]. Since q_i and q_j will generally not be adjacent in real space, we must then perform the following transfer sequence:

1. pick up one atom with an optical tweezer (corresponding to the *Bring Home Water* problem in Fig. 3.3),
2. subsequently transport it to the vicinity of the second atom,
3. perform the $\sqrt{\text{SWAP}}$ protocol, and
4. transport the now-entangled atom back to its initial position to restore the register configuration for further computations.

Apart from increasing gate duration, this is numerically complicated by the fact that the full Hilbert space of two indistinguishable particles is $D_{\mathcal{H}} = (D_{\text{external}} \times D_{\text{internal}})^2$ -dimensional where $D_{\text{external}} = D_x D_y D_z$. Individual typical values are $D_{x/y/z} \sim \mathcal{O}(10^1) - \mathcal{O}(10^2)$ and this means that D_{external} is huge even for crude spatial discretizations as discussed in Fig. 2.4.

To ameliorate numerical treatments we therefore typically reduce the spin and space degrees of freedom by different means and to various extents of approximation as described in Sec. 2.4. Such an analysis is detailed and concretely estimated e.g. for the studied problem is shown in Fig. 6.6 and it is found that there is a motional leakage into the y direction because the potential⁵ is not fully separable. Similar

⁵ The nonseparability incurred by the interaction is small compared and the one dimensional assessment can be made to first order on the basis of single particles. Independent, tight confinement is provided in the z direction.

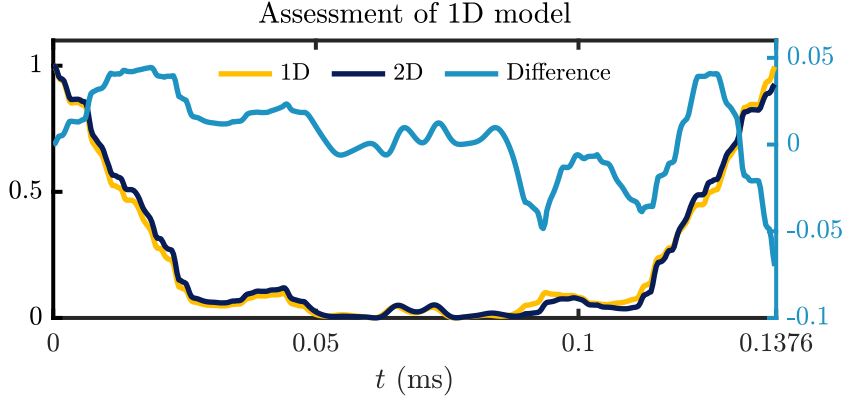


Figure 6.6: Adapted from Ref. [2]. Instantaneous fidelities with the initial state when propagating a single particle (initially in the ground state of the left well) with $\hat{U}_{\sqrt{\text{SWAP}}}^*$ from Fig. 6.5. Results are shown in one dimension $F_{1D} = |\langle \psi_{\text{ini}}(x) | \psi(x, t) \rangle|^2$ and in two dimensions $F_{2D} = |\langle \psi_{\text{ini}}(x, y) | \psi(x, y, t) \rangle|^2$ (left axis). The difference between the two fidelity trajectories (right axis) roughly corresponds to the motional leakage into the y direction.

effects can be expected for an optical tweezer potential because it is also not fully separable. It is remarked that these effects are naturally exaggerated in the highly diabatic limit which is the regime where optimal control is typically operating. However, it would not be hard to introduce new terms in the cost functional that counteract the unwanted behavior. For example, imposing either a maximally allowed energy spread or a smooth penalty hereto would in an optical tweezer geometry mean that the wave packet is always localized near the trap center were the harmonic separability approximation is very good. This would of course incur slower implementation speeds as high energy spreads lead to shorter dynamical time scales, $\Delta t \sim \hbar / \Delta \hat{H}$, as discussed in Sec. 3.8. If this and similar countermeasures can sufficiently reduce the modeling error could be a topic for another work. It is pointed out that our considerations and conclusions both transcend the particular choice of trapping geometry and are also independent of these modeling concerns.

Inescapably, collision-gates require manipulation of at least one external degree of freedom and therefore at the very minimum $D_{\mathcal{H}} = D_x^2$ when the interactions are isotropic (allowing $D_{\text{internal}} = 1$) and $D_{\mathcal{H}} = (D_x \times 2)^2$ for minimally state-selective potentials and/or scattering anisotropy (such as in AEL proposals). This is scaling is, overall, considerably less attractive than in Rydberg-gates where it is the external coordinates that are assumed fixed and we instead consider the Hilbert space $D_{\mathcal{H}} = D_{\text{internal}}^N$ for N atoms in the register (we now cannot just consider two “vacuous” atoms since Rydberg interactions are long-ranged and extend over several sites). Nevertheless, usually only a few finite $D_{\text{internal}} \sim \mathcal{O}(1)$ levels are needed since transitions other than $|0\rangle \leftrightarrow |1\rangle \leftrightarrow |r\rangle$ (i.e. the qubit-encoding states and Rydberg

state transitions and possibly a few additional states from either manifold) are heavily suppressed and negligible when the driving laser is near-resonant. This allows N to be of somewhat reasonable, yet still limited, size in exact diagonalization but this number can be greatly increased using tensor network descriptions as discussed in Ch. 7.

It is therefore currently much more practical to implement quantum computation and simulation using Rydberg gates. If collision-based architectures may become a competitive alternative to Rydberg-based ones, perhaps due to the AELS developments, remains speculative but unlikely in the near-term.

MANUSCRIPT [2] – TIME-OPTIMAL CONTROL OF
COLLISIONAL $\sqrt{\text{SWAP}}$ GATES IN ULTRACOLD ATOMIC
SYSTEMS

Published in Physical Review A.

13 pages.

AUTHOR CONTRIBUTIONS

I performed the analytical and numerical studies, and analysis and interpretation of the presented results. I authored the manuscript and figures. Jens Jakob Sørensen, Jacob Sherson, and Klaus Mølmer participated in scientific discussions and provided scientific feedback and corrections to the manuscript.

QUANTUM mechanics rely fundamentally on superposition and entanglement properties. This is, for instance, what allows quantum computing to achieve advantages over classical computing: with each qubit one gains an exponential increase in computational resources. Due to its inherent expressiveness, however, this also implies that exactly simulating quantum systems with classical resources is exponentially demanding [16]. Here we explore one possible way to circumvent these issues by using describing the quantum system with *matrix product states* and apply exact gradient optimal control to a paradigmatic, yet challenging many-body problem within this ansatz.

The chapter is structured as follows.

Section 7.1 sets the stage and introduces the *curse of dimensionality*.

Section 7.2 details with a sigh of relief that the curse can be broken since physically relevant states only span a tiny part of the vast Hilbert space. This insight is powered by entanglement entropy area scaling laws.

Section 7.3 introduces and details certain elements of the matrix product state tensor network ansatz capable of providing an exceedingly efficient (subexponential) description of low-entanglement states.

Section 7.4 presents a tensor network time evolution algorithm based on control diagonal Trotterized propagators, enabling the exact derivatives of Sec. 4.4.

Section 7.5 provides an introduction to and summary of the manuscript [4], solving the Bose-Hubbard problem for problem parameters far beyond the reach of exact diagonalization.

Section 7.6 quantitatively details the manuscript and its results. Mapping out fidelities $F = 0.99 - 0.9999$ and corresponding minimal duration estimates, optimal control strategies are identified and the resulting dynamics are analyzed.

Section 7.7 highlights pertinent perspectives to other contemporary developments.

7.1 THE CURSE OF DIMENSIONALITY

Suppose we have an arrangement of L sites (either in real or abstract space) and let $|j_i\rangle$ denote a local basis on the i 'th site ($i =$

$1, \dots, L$) with a d_i -dimensional local Hilbert space $\mathcal{H}_{\text{site } i}$. Each local basis can in principle be different, but from hereon we consider the case where $d_i = d$ with all sites having identical (time-independent) numbered local states $|j_i\rangle \in \{|1\rangle, \dots, |d\rangle\}$. These are Fock states of second quantization cf. Sec. 2.3. For example, for a system of spin-1/2 particles we could choose $\hat{\sigma}_z$ -eigenstates as the local basis $\{|1\rangle, |2\rangle\} = \{|\uparrow\rangle, |\downarrow\rangle\}$ with $d = 2$ and for a Bose-Hubbard system with N_{part} particles we could choose occupational $\hat{n}_i$ -eigenstates¹ $\{|1\rangle, \dots, |d\rangle\} = \{|0\rangle, \dots, |N_{\text{part}}\rangle\}$ with $d = N_{\text{part}} + 1$.

The full Hilbert space is then formed by aggregating L copies of $\mathcal{H}_{\text{site}}$, i.e. $\mathcal{H} = \mathcal{H}_{\text{site}}^{\otimes L}$. A natural basis for $\mathcal{H}$ is clearly

$$\{|j\rangle\} = \left\{ |j_1, j_2, \dots, j_L\rangle \mid j_i \in \{1, \dots, d\}, \forall i \in \{1, \dots, L\} \right\} \quad (7.1)$$

and expanding an arbitrary quantum state² hereon yields

$$|\psi\rangle = \sum_j c_j |j\rangle = \sum_j c_{j_1, \dots, j_L} |j_1, \dots, j_L\rangle = \sum_{j_1}^d \cdots \sum_{j_L}^d c_{j_1, \dots, j_L} |j_1, \dots, j_L\rangle. \quad (7.2)$$

Thus, as also shown in Fig. 7.1 for the mentioned examples above, the number of state coefficients required for completely describing just a single, generic state undergoes a combinatorial explosion as L increases (compare also Fig. 2.4). Associated computations and algorithms to do anything “useful” scale in powers of these fundamentally detrimental requirements.

This is the much lamented *curse of dimensionality* of exact diagonalization and its feasibility for studying quantum systems in e.g. the many-body regime (and similarly expansive Hilbert spaces) is severely limited. This is especially worrisome since a large variety of interesting many-body phenomena such as phase diagrams and transitions [149] are inherently bulk effects in the thermodynamic limit. In rough estimation, exact diagonalization can feasibly reach low tens of L in spin-1/2 systems [312–314] while for Bose-Hubbard systems only $L = N_{\text{sites}} = N_{\text{part}} = 10 - 13$ [315, 316] and attaining the higher echelons requires increasingly involved methodologies [315, 317] for relatively small gains.

¹ Recall from Sec. 2.5 that these occupational states are in reference to site-localized Wannier wave functions.

² We will assume a fixed-particle number for e.g. the Bose-Hubbard model such that the expansion is only on states consistent with this requirement. More generally, if an operator $\hat{A}$ commutes with the Hamiltonian, $[\hat{H}(t), \hat{A}] = 0$, then its quantum numbers a defined by $\hat{A} |a\rangle = a |a\rangle$ are constants of motion. Basis states can then be grouped according to their value of a and dynamics between these groupings are prohibited and thus separable. In modeling a situation for some particular a' , it is therefore sufficient to only include basis states consistent with this “good” quantum number in the expansion. One example was $\hat{N} = \sum_i \hat{n}_i$ in the Bose-Hubbard model and another could be the total spin projection $\hat{S}^z = \sum_i \hat{s}_i^z$ in certain spin Hamiltonians.

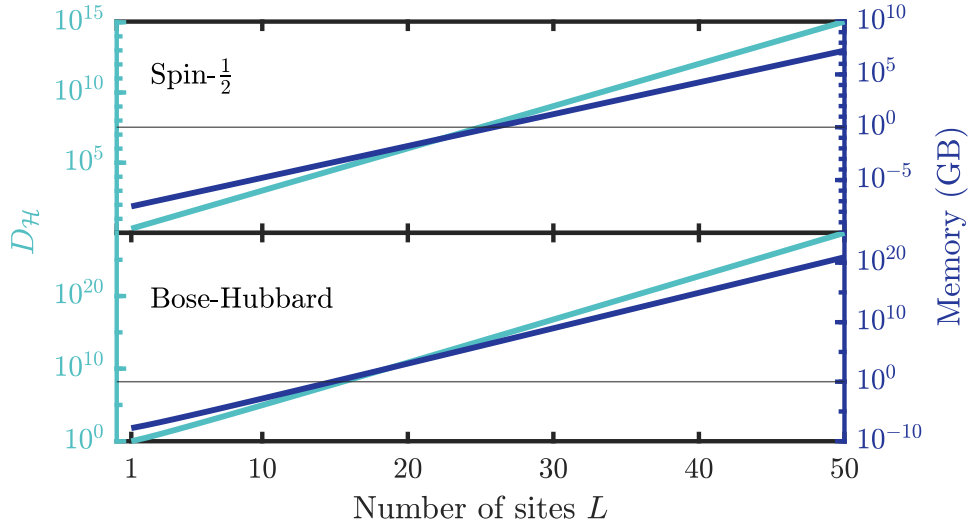


Figure 7.1: Example scaling of Hilbert space and the required computer memory to store a single, generic state vector using double-precision floating point numbers. Here $D_{\mathcal{H}} = 2^L$ and $D_{\mathcal{H}} = (N_{\text{sites}} + N_{\text{part}} - 1)! / (N_{\text{part}}!(N_{\text{sites}} - 1)!)$ for spin-1/2 and the Bose-Hubbard model (taking unit filling $N_{\text{part}} = N_{\text{sites}} = L$), respectively.

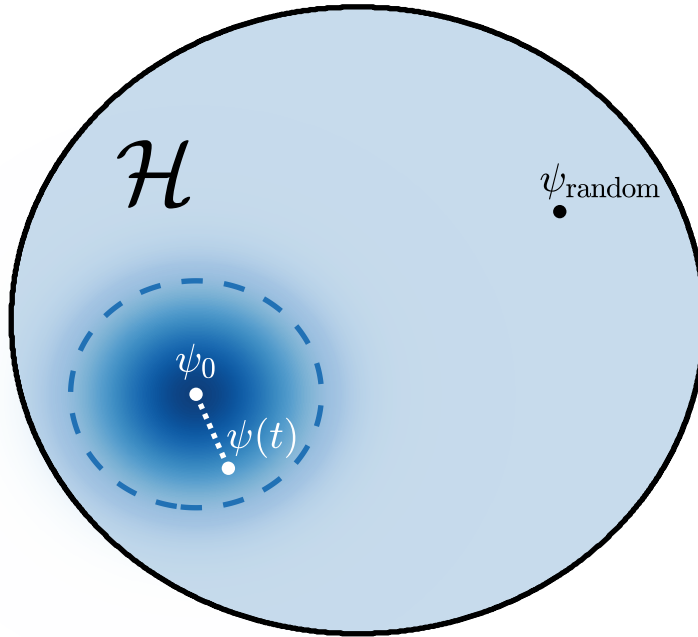


Figure 7.2: The convenient illusion of Hilbert space $\mathcal{H}$. The physically relevant states, i.e. those reachable from for example ground states in finite polynomial time $|\psi(t)\rangle = \hat{U}(T;0)|\psi_0\rangle$, reside in a tiny corner of the full space (the illustration is not to scale). Tensor network ansätze are well-equipped for targeting exactly these low-entanglement states that obey an area scaling law. Reaching highly entangled states is prohibited by exponential evolution times.

In attempting to map out e.g. high- T^c superconductivity in solid state systems of sizes on the order of Avogadro's number, $L \leq 50$ atoms can scarcely be considered a true thermodynamic limit. More relevant to the present optimal control context, these numbers are substantially hampered even further as we must solve the equations of motion to an excessive degree during optimization. As a preview of Ch. 8, we may divide the difficulty of solving a particular control problem into (orthogonal) axes of numerical complexity and inherent topographical complexity. Considering presently just the former, it is constituted by the computational runtime of (1) the optimal control algorithm, and (2) simulating the quantum system and its dynamics. Minimizing runtime (1) while retaining locally efficient and reliable landscape maneuvering was discussed in great detail in Ch. 4. To efficiently and optimally control many-body systems we must therefore also devise methodologies to minimize runtime (2).

7.2 BREAKING THE CURSE

What may seemingly appear an impossibly vast Hilbert space is, in a certain sense, a convenient illusion [318, 319]. Qualitatively, not all expansions and terms $c_j |j\rangle$ in Eq. (7.2) are equally necessary to consider. This may appear somewhat obvious, but the extent to which this is true is as surprising as it is subtle.

Let us consider the “worst imaginable” states in terms of nonzero coefficients and thus required storage memory, for example the equal superposition state $c_j = (\sqrt{D_{\mathcal{H}}})^{-1}$ or a randomly generated (properly normalized) state $c_j \sim \text{uniform}(0, 1)$. A precise quantitative characteristic of both is that they correspond to highly excited and entangled states (for appreciable $L, d, D_{\mathcal{H}}$). The saving grace is, then, that it can be formally shown [318, 320] under reasonable assumptions³ that these “worst imaginable” states are also “improbable” states in these sense that they take exponential amounts of evolution time to reach from what we may consider “physically relevant” states, e.g., ground and low/modestly excited and entangled states. In other words, the physically relevant states occupy only a tiny fraction of the full Hilbert space as conceptually visualized in Fig. 7.2.

A suitable measure for entanglement is the so-called *entanglement entropy* [321]. Consider some bipartitioning (defined by some bounding region) of the full system and the entanglement entropy between the two subsystems. Rather than the exact value for a given state, the general scaling behavior of the entanglement entropy when considering increasing region sizes⁴ turns out to be of importance. It can be

³ Sufficiently local and bounded interaction elements for spectrally gapped Hamiltonians (or alternatively approximately/less rigorously for ungapped ones).

⁴ Visualize, e.g., a large (possibly infinite) system with an inflating electrodynamic-like Gaussian surface. This surface defines a boundary region and an interior volume.

concretely shown that the entanglement entropy of physically relevant states are formally described in terms of so-called “area” scaling laws. [56, 318, 319, 321–326]. That is, the entropy scales with the area of the region boundary rather than its volume. The “dimensionality” of the area itself also depends on the type of system (usually correlated with the actual spatial dimensionality/connectivity) at hand. States fulfilling such area laws will continue to do so upon time evolution⁵.

It is of course not enough to know that these tiny corners of Hilbert space exist — we must also be able to efficiently target and mathematically describe them with techniques distinct from exact diagonalization. What turns out to be a natural candidate in this regard are so-called *tensor networks*. We will explicitly see certain examples where low entanglement states are well-suited for such descriptions.

7.3 TENSOR NETWORKS

Tensor networks are a broad class of mathematical *anzätze* that has spawned a rich literature in physics [56, 275, 319, 324, 325, 328–330] since they have turned out to be very useful for describing quantum states. In fact, the adopted techniques and developments provide a paradigmatic tonal shift in how to describe, perform, and think about quantum mechanical computations. That is, the tensor network structure necessitates a new framework for calculating states, overlaps, operators, expectation values, diagonalization, time evolution, and so on, with subexponential/polynomial resources (both time and memory).

Tensor networks remain an active and advanced research area [331–333] and it is well beyond the scope and competency of this thesis to give an in-depth review. We will instead take a pragmatic stance that introduces the base conceptual ideas and simplest tensor network known as *matrix product states* (or *tensor trains*) which is the appropriate type of network for describing one-dimensional systems. For the tensor computations in Ref. [4] we use the C++ library ITENSOR [334] and instead focus our programming efforts primarily on our optimal control competencies. For a more complete and technical introduction to matrix product states I recommend Ref. [275] and much of the following was learned and inspired from there. This reference also establishes the historical and mathematical link to Density Matrix Renormalization Group (DMRG) — the nomenclature of which still lingers on today, for example in diagonalization and certain time evolution algorithms. I personally find this somewhat confusing on occasion.

⁵ The prefactor is, however, expected to grow exponentially with time which ultimately limits very long time scales [327] for the algorithms described in the following, at least if the dynamics is not properly constrained.

Tensor network notation and computations are somewhat cumbersome when written in usual mathematical expressions as evidenced shortly. It turns out, however, that these are very amenable to beautiful and simplified pictorial depictions and interpretations (playing a similar role as quantum circuit diagrams). We will only see one example here in discussing a time evolution algorithm, but these will otherwise not be covered here in detail for the sake of brevity.

7.3.1 Tensors

A tensor $T_{i,j,\dots,q}$ is simply an indexed structure of data, a way of storing numbers in our case. The rank of a tensor is equal to the number of indices. For example, T is a rank 0 tensor (scalar), T_i is a rank 1 tensor (vector), $T_{i,j}$ is a rank 2 tensor (matrix), $T_{i,j,k}$ is a rank 3 tensor (a matrix with “depth”), and so on. Each index is of a particular size/dimension, e.g. $i = 1, 2, \dots, d_i$, and the indices of tensors are generally both different from each other and of different dimension.

Tensors with at least one common index can be *contracted* which is a generalization of matrix multiplication

$$C_{i,k} = \sum_j A_{i,j} B_{j,k} \equiv A_{i,j} B_{j,k}, \quad (7.3)$$

where repeated indices imply (Einstein) summation. On an abstract level the indices are uniquely identified by their name and so their order and notation does not matter (it may be relevant in software implementations). Thus $T_{i,j,k} = T_{i,k,j} = T_{j,k}^i$ and so on is the same thing, but different notations can be used to emphasize conceptual differences as we will see shortly.

An important insight is that the data contained in a tensor does not depend on the structure. For example, the information in the rank 3 tensor $T_{i,j,k}$ could just as well have been stored as a vector $T_{(i,j,k)}$ or a matrix $T_{i,(j,k)}$ or $T_{(i,j),k}$. Here, e.g., (i,k) is called a *fused* index [56] for apparent reasons and the inverse process is called *index splitting*. There is no loss of data in these processes, just a rearrangement of the underlying storage structure⁶. However, upon fusing indices of

⁶ As a simple example, consider the reshape functions in MATLAB and Python. Fusing indices introduces a new (linear) index k that somehow maps to the fused ones, $k \xleftrightarrow{\text{map}} (i,j)$, and oppositely for splitting. The matrix $T = [5, 6; 7, 8]$ is a rank 2 tensor $T_{i,j}$ — reshaping this matrix to $T \rightarrow [5, 6, 7, 8]$ is equivalent to fusing the indices such that we now have a rank 1 tensor $T_k = T_{(i,j)}$ with the following index mappings

$$\begin{aligned} k = 1 &\leftrightarrow (1, 1) = (i, j), \\ k = 2 &\leftrightarrow (1, 2) = (i, j), \\ k = 3 &\leftrightarrow (2, 1) = (i, j), \\ k = 4 &\leftrightarrow (2, 2) = (i, j). \end{aligned}$$

The data is the same regardless of the chosen structure.

a rank L tensor $T_{i_1, \dots, i_L}$ such that the data is stored as a matrix (rank 2 tensor), e.g. $T_{(i_1, \dots, i_l)(i_{l+1}, \dots, i_L)}$, one can now invoke the expansive and eminently useful machinery of matrix algebra. Upon completing any desired operation there, yielding e.g. a modified $\tilde{T}_{(i_1, \dots, i_l)(i_{l+1}, \dots, i_L)}$, one can then subsequently split the indices back into the original format, $\tilde{T}_{i_1, \dots, i_L}$.

7.3.2 Compressing Data

Singular Value Decomposition

A particularly strong tool in the following is the [SVD](#) of any given matrix $A \in \mathbb{C}^{N \times M}$. The decomposition⁷ reads

$$\underbrace{A}_{N \times M} = \underbrace{U}_{N \times K} \times \underbrace{S}_{K \times K} \times \underbrace{V^\dagger}_{K \times M}, \quad \text{where } K = \min(N, M), \quad (7.4a)$$

$$A_{i,j} = \sum_{a=1}^K U_{i,a} S_{a,a} V_{j,a}^* = \sum_{a=1}^K U_{i,a} S_{a,a} (V^\dagger)_{a,j}. \quad (7.4b)$$

in full and element-wise notation, respectively. The three matrices U, S, V , have very nice properties [275].

1. U has orthonormal columns (left singular vectors), $U^\dagger U = I$.
2. S is diagonal with nonnegative entries $S_{a,a} = s_a$ called singular values. These are usually ordered in descending fashion,

$$s_1 \geq s_2 \geq \dots s_r > s_{r+1} = s_{r+2} = \dots = s_K = 0. \quad (7.5)$$

The *Schmidt rank*, r , of A is the number of nonzero singular values.

3. $V^\dagger$ has orthonormal rows (right singular vectors), $V^\dagger V = I$.

The [SVD](#) is useful for many things, for example compression of data as follows.

The decomposition matrices can be truncated since only the first r (i.e. up to the Schmidt rank) singular values are nonzero and contributes to the overall reconstruction of the original matrix A . Thus, the decomposition matrices can be truncated by removing rows and columns with indices greater than r in S and removing corresponding columns in U and rows in $V^\dagger$. This shrinks the dimension of e.g. S from $K \rightarrow r$, but still allows exact reproduction of the original matrix. Of course if all $s_a > 0$ then $r = K$ and no *lossless* compression is possible. Nevertheless, because of the nonnegative ordering, we can always define some index cutoff threshold r' and discard all following singular values that are usually smaller — a similar thing can be done

⁷ More specifically, we discuss the so-called reduced [SVD](#)— the “full” [SVD](#) consists in padding the U matrix such that it is square and still orthonormal.

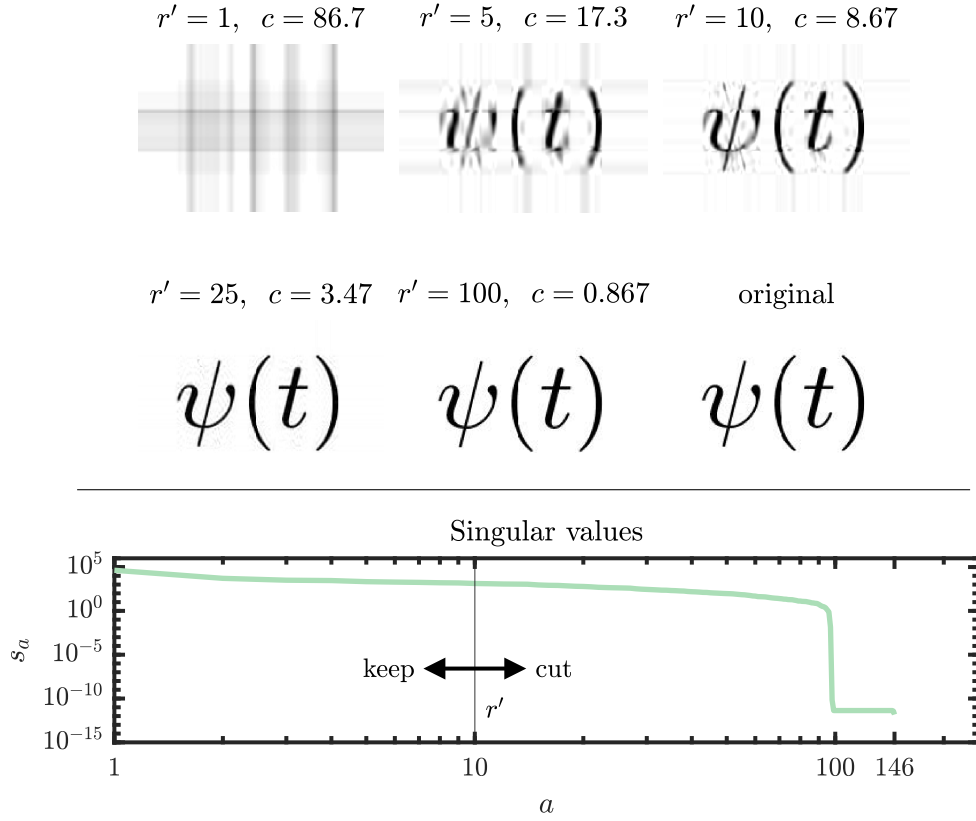


Figure 7.3: Compression of a $A \in \mathbb{R}^{215 \times 146}$ matrix corresponding to a grayscale image. The original matrix (lower right in upper panels) is reconstructed, $A \approx USV^\dagger$, for different singular value index cutoffs r' in the three decomposition matrices. For each r' the compression factor c is shown, i.e. the relative reduction in storing the full original A and the matrices $U, S, V^\dagger$. The singular values of the [SVD](#) is shown in the lower panel. Here the cutoff at $r' = 10$ is exemplified.

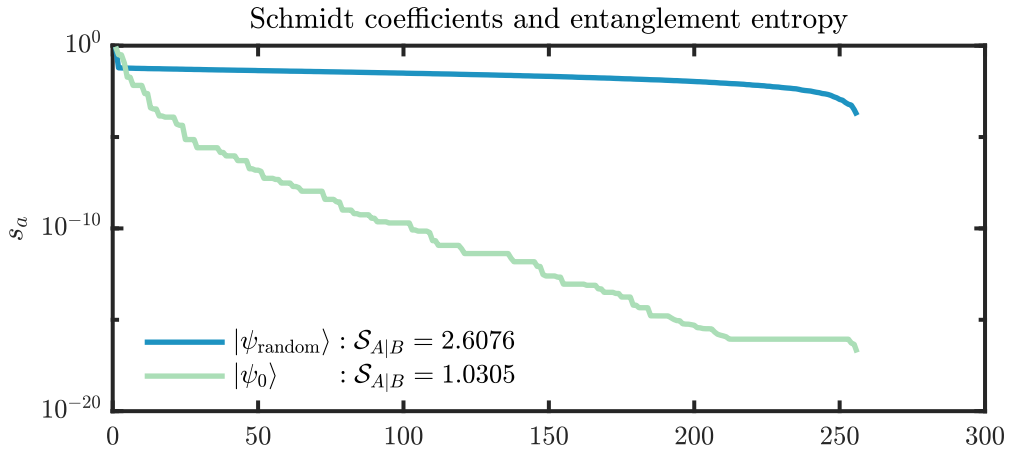


Figure 7.4: Singular values (Schmidt coefficients) for a spin-1/2 chain ($d = 2, L = 16$) for a randomly generated state and the ground state of $\hat{H} = \sum_i \hat{S}_x^i \hat{S}_x^{i+1} + \hat{S}_z^i \hat{S}_z^{i+1}$ where $\hat{S}_i$ are spin operators. The entanglement entropy $\mathcal{S}_{A|B}$ is measured across the centered bipartition $A|B = (1, \dots, L/2)|(L/2 + 1, \dots, L)$.

for some nonzero threshold s_{thresh} . This shrinks the dimension of S from $K \rightarrow r' < r$ and similarly for U and $V^\dagger$. Although this introduces data loss, this is in fact the optimal approximation of A with respect the Frobenius norm [275]. Figure 7.3 shows an example of SVD compressing a matrix corresponding to a grayscale image. By visual inspection it is evident that even a highly compressed $r' \ll K = 146$ approximation retains most of the structure and information of the original matrix. We also find, unsurprisingly, that there is a compression/reproducibility trade-off and for large enough r' there is no gain in the decomposition ($c < 1$). The original matrix is hugely influential on the necessary threshold r' or s_{thresh} for a sufficiently good approximation and this is largely determined by the singular value rate of decay. In fact, this example matrix is not particularly well-behaved in this regard since it has relatively many high-value singular values and a slow decay up until $a \approx 100$ where a significant drop is finally found. Conversely, the ideal situation occurs when there are only relatively few large singular values that decays e.g. exponentially which can yield a nearly lossless, highly compressed data storage.

Schmidt Decomposition

Now let us connect some of these ideas to a quantum mechanical setting. Consider a pure state on the joint Hilbert space of bipartite subsystems A and B with some orthonormal bases $\{|i\rangle_A\}$ and $\{|j\rangle_B\}$, respectively. We expand on the joint basis and think of the coefficients $c_{i,j}$ as a matrix (rank 2 tensor) on which we can perform an SVD and subsequently distribute the sums

$$\begin{aligned}
 |\psi\rangle &= \sum_{i,j} \underbrace{c_{i,j}}_{\text{SVD}} |i\rangle_A |j\rangle_B = \sum_{i,j} \left(\sum_a^K U_{i,a} S_{a,a} V_{j,a}^* \right) |i\rangle_A |j\rangle_B \\
 &= \sum_a^K s_a \underbrace{\left(\sum_i U_{i,a} |i\rangle_A \right)}_{|a\rangle_A} \underbrace{\left(\sum_j V_{j,a}^* |j\rangle_B \right)}_{|a\rangle_B} \\
 &= \sum_a^K s_a |a\rangle_A |a\rangle_B = \sum_a^r s_a |a\rangle_A |a\rangle_B, \tag{7.6}
 \end{aligned}$$

where of course $\sum_a s_a^2 = 1$ and the last truncation equality is due to the definition of the Schmidt rank. Due to U and $V^\dagger$ orthogonality properties, the new states $\{|a\rangle_A\}$ and $\{|a\rangle_B\}$ also orthonormal [335] and can form a basis for the joint system. What we have discovered here is the *Schmidt decomposition*. The Schmidt decomposition is often introduced separately (s_a are also called *Schmidt coefficients*), but from this small calculation its equivalence to the SVD is obvious. The central point is that we can also approximate and compress *quantum states* $|\psi'\rangle \approx |\psi\rangle$ in full colloquial analogy to Fig. 7.3, hence the choice of original matrix there. That is, by decomposing $c_{i,j} = USV^\dagger$ we can

truncate the singular value matrix S by $K \rightarrow r' < r$, which does require re-normalization, and similar for $U, V^\dagger$. The question is then: for which quantum states $|\psi\rangle$ can we attain an efficient representation, i.e. a high compression factor c ?

To make headways in answering this question, we now recall that states that cannot be written as single product are said to be entangled. It follows that $r = 1$ corresponds to an unentangled state while $r > 1$ corresponds to entangled states. In this sense the singular values s_a measure the entanglement between the A and B partitions. Specifically, the von Neumann entanglement entropy between A and B , assuming proper normalization, reads

$$\mathcal{S}_{A|B}(\psi) = - \sum_a^r s_a^2 \log_2 s_a^2. \quad (7.7)$$

Finally, we can conclude that physical quantum states that are *well-suited* for the above-described approximation machinery are the ones characterized by low entanglement — i.e. exactly those encompassing the tiny corner of Hilbert space discussed in relation to Fig. 7.2 in Sec. 7.2. As a concrete numerical example, Fig. 7.4 shows the singular values and entanglement entropy for a randomly generated state $|\psi_{\text{random}}\rangle$ and a Hamiltonian ground state $|\psi_0\rangle$. The singular values decay rapidly for the ground state and can therefore efficiently and compactly be described by *matrix product states* as elucidated in the following section.

The Schmidt decomposition is by itself, however, not immediately of practical use for computation as it describes a bipartitioning of the system whereas we are interested in considering an *ab initio* expansion in terms of the local basis $|j\rangle$ by Eq. (7.2). In the following we will therefore instead start from the $|j\rangle$ expansion and this will lead us to the matrix product state description. Then, we will finally work our way back to the Schmidt decomposition so that we can also measure the entanglement entropy if so desired.

7.3.3 Matrix Product States

We are now in a position to introduce the Matrix Product State (MPS). We assume that the system is a one-dimensional chain and it turns out that an MPS description of such systems is appropriate because the boundary area (in the sense of entropy scaling laws) formed by taking a cut through the chain is constant, i.e. independent on the number of sites [321].

To proceed, we consider the state coefficients $c_{j_1, j_2, \dots, j_L}$ of an arbitrary state in Eq. (7.2) as a tensor of rank L . The indices j_i are called *physical indices* and as before these have dimension d . The main conceptual idea is to successively apply SVDs as in the first line of Eq. (7.6).

Depending on from which side of the chain we start, however, we will influence the properties of the produced matrices.

Left-canonical MPS

Here we start from the leftmost site and proceed to the right, $1 \rightarrow L$, and vice versa in the next section. To begin, and since we can store information in whatever format, we fuse indices $(j_2, \dots, j_L)$ to obtain a $(d \times d^{L-1})$ -dimensional matrix which we can now subject to an [SVD](#)

$$\begin{aligned} c_{j_1, j_2, \dots, j_L} &\xrightarrow{\text{fuse}} c_{j_1, (j_2, \dots, j_L)} \xrightarrow{\text{SVD}} \sum_{a_1=1}^{r_1} U_{j_1, a_1} \underbrace{S_{a_1, a_1} (V^\dagger)_{a_1, (j_2, \dots, j_L)}}_{c_{a_1, (j_2, \dots, j_L)}} \\ &\xrightarrow{\text{split}} \sum_{a_1=1}^{r_1} \underbrace{U_{j_1, a_1}}_{A_{1, a_1}^{j_1}} c_{a_1, j_2, \dots, j_L} = \sum_{a_1=1}^{r_1} A_{1, a_1}^{j_1} c_{a_1, j_2, \dots, j_L}, \end{aligned} \quad (7.8)$$

where r_1 is again the Schmidt rank for the decomposition and we contracted S and $V^\dagger$ by Eq. (7.3) to form a new coefficient tensor. The [SVD](#) introduced an *auxiliary index/bond index* a_1 . The size of a_1 is equal to the Schmidt rank r_1 which is therefore also referred to as the bond- or auxiliary-dimension.

We also utilized our notational freedom to help with conceptual convenience: in defining $A_{1, a_1}^{j_1} = U_{j_1, a_1}$, the rank 3 tensor $A_{1, a_1}^{j_1}$ can be thought of as a collection of d row vectors⁸ or a matrix of depth d with $\text{height} \times \text{width} = (1 \times r_1)$. Note that the physical index j_1 now *only* shows up in $A_{1, a_1}^{j_1}$.

We then continue with fusing indices of $c_{a_1, j_2, \dots, j_L}$ into a matrix of size $(r_1 \cdot d \times d^{L-2})$ and perform another [SVD](#)

$$\begin{aligned} \dots &= \sum_{a_1}^{r_1} A_{1, a_1}^{j_1} c_{a_1, j_2, \dots, j_L} \xrightarrow{\text{fuse}} \sum_{a_1}^{r_1} A_{1, a_1}^{j_1} c_{(a_1, j_2), (j_3, \dots, j_L)} \\ &\xrightarrow{\text{SVD}} \sum_{a_1}^{r_1} \sum_{a_2}^{r_2} A_{1, a_1}^{j_1} \left(U_{(a_1, j_2), a_2} \underbrace{S_{a_2, a_2} (V^\dagger)_{a_2, (j_3, \dots, j_L)}}_{c_{a_2, (j_3, \dots, j_L)}} \right) \\ &= \sum_{a_1}^{r_1} \sum_{a_2}^{r_2} A_{1, a_1}^{j_1} U_{(a_1, j_2), a_2} c_{a_2, (j_3, \dots, j_L)} \\ &\xrightarrow{\text{split}} \sum_{a_1}^{r_1} \sum_{a_2}^{r_2} A_{1, a_1}^{j_1} A_{a_1, a_2}^{j_2} c_{a_2, j_3, \dots, j_L}, \end{aligned} \quad (7.9)$$

where we contracted S and $V^\dagger$ into a new coefficient tensor and use our notational freedom to write $A_{a_1, a_2}^{j_2} = U_{a_1, j_2, a_2}$. The bond index a_2 has dimension r_2 . We can think of this rank 3 tensor as either a matrix of depth d with $\text{height} \times \text{width} = (r_1 \times r_2)$ or a collection of d matrices

⁸ The subscripted $_1$ on $A_{1, a_1}^{j_1}$ is a dummy index to remind this.

of the same size. Again the physical index j_2 now only appears on the tensor for site 2, $A_{a_1,a_2}^{j_2}$, but it shares the bond index a_1 with the tensor for site 1, $A_{1,a_1}^{j_1}$. These two tensors are therefore valid for contraction. On the other hand, $A_{1,a_1}^{j_1}$ and $c_{a_2,j_3,\dots,j_L}$ *cannot* be directly contracted because no indices are shared.

We continue this **SVD** pattern, each time stripping a physical index from the temporary c tensor. If we do this L times we end up with

$$c_{j_1,j_2,\dots,j_L} = \sum_{a_1}^{r_1} \sum_{a_2}^{r_2} \cdots \sum_{a_L}^{r_L} A_{1,a_1}^{j_1} A_{a_1,a_2}^{j_2} A_{a_2,a_3}^{j_3} \cdots A_{a_{L-1},1}^{j_L}, \quad (7.10)$$

where we introduced a dummy label 1 on the last tensor to remember this rank 3 tensor as a matrix of depth d with height/width $(r_L \times 1)$. Thus we have split the monolithic rank L coefficient tensor into L rank 3 tensors. As a generalized statement of the possible contractions above, note that the tensor at site i only “talks” to its nearest-neighbors through the corresponding common bond index a_i and a_{i+1} . The index connectivity of the tensors form a network, hence the apt name *tensor network*.

Suppose we fix the index vector $\mathbf{j}$, corresponding to a particular coefficient — then we simply have a $L - 1$ matrix multiplications or contractions as defined in Eq. (7.3) which turn into a scalar!

$$c_{j_1,j_2,\dots,j_L} = \mathbf{A}^{j_1} \mathbf{A}^{j_2} \mathbf{A}^{j_3} \cdots \mathbf{A}^{j_L}, \quad (7.11a)$$

$$\Rightarrow |\psi\rangle = \sum_{\mathbf{j}} \mathbf{A}^{j_1} \mathbf{A}^{j_2} \mathbf{A}^{j_3} \cdots \mathbf{A}^{j_L} |j_1, j_2, \dots, j_L\rangle. \quad (7.11b)$$

That is, we describe our state with products of matrices (here highlighted in bold for emphasis) — hence a matrix product state. An **MPS** is, in principle, an *exact* ansatz for the state coefficients $c_{j_1,j_2,\dots,j_L}$ if keep all nonzero singular values. In practice one employs the very same **SVD** truncation/compression schemes described in Sec. 7.3.2 (or similar) — otherwise we just obtain a more expensive way of describing the state than with the original coefficients. This exactly mirrors the image compression example of Fig. 7.3 and there is similarly no *a priori* guarantee that the exponential resource scaling is tamed by using **MPS** for any other states than the physically relevant ones.

Can we obtain a Schmidt decomposition from this? Not quite. Note first that at each **SVD** we have $U^\dagger U = I$ and from this it can be shown that

$$\begin{aligned} \delta_{a_l, a'_l} &= (U^\dagger U)_{a_l, a'_l} = \sum_{a_{l-1}, j_l} (U^\dagger)_{a_l, (a_{l-1}, j_l)} U_{(a_{l-1}, j_l), a'_l} = \sum_{j_l} \left(\sum_{a_{l-1}} (A^{j_l \dagger})_{a_l, a_{l-1}} A_{a_{l-1}, a'_l}^{j_l} \right) \\ &= \sum_{j_l} \left(A^{j_l \dagger} A^{j_l} \right)_{a_l, a'_l} \Rightarrow I = \sum_{j_l} A^{j_l \dagger} A^{j_l}. \end{aligned} \quad (7.12)$$

The set of matrices A^{j_l} at site l are then called *left-normalized* and an MPS consisting of just such matrices are called *left-canonical*. Now suppose we divide the chain at some site l into the following states (intermediate a_i have been contracted)

$$|a_l\rangle_A = \sum_{j_1, \dots, j_l} (A^{j_1} A^{j_2} \dots A^{j_l})_{1, a_l} |j_1, j_2, \dots, j_l\rangle, \quad (7.13)$$

$$|a_l\rangle_B = \sum_{j_{l+1}, \dots, j_L} (A^{j_{l+1}} A^{j_{l+2}} \dots A^{j_L})_{a_l, 1} |j_{l+1}, j_{l+2}, \dots, j_L\rangle, \quad (7.14)$$

$$\Rightarrow |\psi\rangle = \sum_{a_l} |a_l\rangle_A |a_l\rangle_B. \quad (7.15)$$

This is only *almost* a Schmidt decomposition because

$$\begin{aligned} {}_A \langle a'_l | a_l \rangle_A &= \left(\sum_{j''_1, \dots, j''_l} (A^{j''_1} A^{j''_2} \dots A^{j''_l})_{1, a'_l} |j''_1, j''_2, \dots, j''_l\rangle \right)^\dagger \\ &\quad \sum_{j_1, \dots, j_l} (A^{j_1} A^{j_2} \dots A^{j_l})_{1, a_l} |j_1, j_2, \dots, j_l\rangle \\ &= \sum_{j_1, \dots, j_l} (A^{j_1 \dagger} \dots A^{j_l \dagger} A^{j_1})_{1, a'_l} (A^{j_1} A^{j_2} \dots A^{j_l})_{1, a_l} \\ &= \sum_{j_1, \dots, j_l} (A^{j_1 \dagger} \dots A^{j_l \dagger} A^{j_1 \dagger} A^{j_1} A^{j_2} \dots A^{j_l})_{a'_l, a_l} \\ &= \sum_{j_2, \dots, j_l} (A^{j_1 \dagger} \dots A^{j_l \dagger} \underbrace{\left(\sum_{j_1} A^{j_1 \dagger} A^{j_1} \right)}_I A^{j_2} \dots A^{j_l})_{a'_l, a_l} \\ &= \dots = I_{a'_l, a_l} = \delta_{a'_l, a_l}. \end{aligned} \quad (7.16)$$

but the same is not generally true for

$$\begin{aligned} {}_B \langle a'_l | a_l \rangle_B &= \sum_{j_{l+1}, \dots, j_L} (A^{j_{l+1}} \dots A^{j_L} A^{j_L \dagger} \dots A^{j_{l+1}})_{a'_l, a_l} \\ &= \sum_{j_{l+1}, \dots, j_{L-1}} (A^{j_{l+1}} \dots A^{j_{L-1}} \underbrace{\left(\sum_{j_L} A^{j_L} A^{j_L \dagger} \right)}_{\neq I} A^{j_{L-1} \dagger} \dots A^{j_{l+1}})_{a'_l, a_l} \\ &\neq \delta_{a'_l, a_l}. \end{aligned} \quad (7.17)$$

So only the $|a_l\rangle_A$'s are orthonormal while $|a_l\rangle_B$'s are not and their joint state cannot be a Schmidt decomposition [275].

Right-canonical MPS

Here we start from the rightmost site and proceed to the left, doing exactly what we did in the previous section. The only exception is that we would now for each site have matrices $B^{j_l}_{a_{l-1}, a_l} = (V^\dagger)_{(a_{l-1} j_l), a_l}$ and the products of U and S contracted into the new coefficient tensor. This yields an MPS of the form

$$c_{j_1, j_2, \dots, j_L} = B^{j_1} B^{j_2} B^{j_3} \dots B^{j_L} \quad (7.18a)$$

$$\Rightarrow |\psi\rangle = \sum_j B^{j_1} B^{j_2} B^{j_3} \dots B^{j_L} |j_1, j_2, \dots, j_L\rangle. \quad (7.18b)$$

The crucial difference to Eqs. (7.11) is that the B matrices inherit the unitary properties of $V^\dagger$,

$$I = \sum_{j_l} B^{j_l} B^{j_l \dagger}. \quad (7.19)$$

The B^{j_l} matrices on site l are therefore called *right-normalized* and MPS consisting fully from these are called *right-canonical*. Partitioning into the states $|a_l\rangle_A |a_l\rangle_B$ this time yields only orthonormal $|a_l\rangle_B$, and so these sets cannot jointly describe the Schmidt decomposition either [275].

Mixed-Canonical MPS

Finally, we can mix the normalizations to directly obtain Schmidt decomposition states. For example,

$$\begin{aligned} c_{j_1, j_2, \dots, j_L} &\stackrel{\text{SVD from } 1 \rightarrow l}{=} \sum_{a_l} (A^{j_1} A^{j_2} \dots A^{j_l})_{1, a_l} S_{a_l, a_l} (V^\dagger)_{a_l, (j_{l+1}, \dots, j_L)} \\ &\stackrel{\text{SVD from } l \leftarrow L}{=} \sum_{a_l} (A^{j_1} A^{j_2} \dots A^{j_l})_{1, a_l} S_{a_l, a_l} (B^{j_{l+1}} B^{j_{l+2}} \dots B^{j_L})_{a_l, 1} \\ &= A^{j_1} \dots A^{j_l} S B^{j_{l+1}} \dots B^{j_L}, \end{aligned} \quad (7.20)$$

and fixing j yields a so-called *mixed-canonical* MPS

$$c_{j_1, j_2, \dots, j_L} = A^{j_1} \dots A^{j_l} S B^{j_{l+1}} \dots B^{j_L}, \quad (7.21a)$$

$$\Rightarrow |\psi\rangle = \sum_j A^{j_1} \dots A^{j_l} S B^{j_{l+1}} \dots B^{j_L} |j_1, j_2, \dots, j_L\rangle. \quad (7.21b)$$

This is a Schmidt decomposition [275] as orthonormality is preserved by the set of A and B matrices

$$|a_l\rangle_A = \sum_{j_1, \dots, j_l} (A^{j_1} A^{j_2} \dots A^{j_l})_{1, a_l} |j_1, j_2, \dots, j_l\rangle, \quad (7.22)$$

$$|a_l\rangle_B = \sum_{j_{l+1}, \dots, j_L} (B^{j_{l+1}} B^{j_{l+2}} \dots B^{j_L})_{a_l, 1} |j_{l+1}, j_{l+2}, \dots, j_L\rangle, \quad (7.23)$$

$$\Rightarrow |\psi\rangle = \sum_{a_l} s_{a_l} |a_l\rangle_A |a_l\rangle_B, \quad (7.24)$$

with $s_{a_l} = S_{a_l, a_l}$, i.e. the Schmidt coefficients are the “leftover” SVD singular values from the bond between sites $(l, l+1)$. From this and our earlier discussion on Schmidt decompositions, we realize that the bond-dimension r_l and the bond index a_l itself characterize the entanglement between the particular partitions A and B . Hence, we could either take the $\max r_l$ or $\mathcal{S}_{A|B}(\psi)$ over all possible partitions to characterize the overall entanglement of any given state [324].

In conclusion, the MPS is thus an exceptionally good ansatz for describing states with low/modest entanglement such that the bond dimensions can be sufficiently truncated and the resource scaling with system size is not exponential. We have now convincingly reached the same result — that low entanglement allows an efficient representation — from multiple different perspectives.

Further Remarks

The matrix product states exposition above is only a base introduction showing how and why they can potentially be useful. The primary concern with tensor networks is to operate with subexponential resources. One glaring problem is, then, that in deriving the canonical MPS forms we applied successive SVDs starting from c_j — but we neither can nor want to store c_j in the first place! The derivations are in this sense not prescriptive in the numerical sense, but are rather meant to legitimize the MPS form and explore their inherent properties. In their implementational deployment there are therefore several additional topics we must consider and adhere to. We briefly discuss a non-exhaustive, compactified list of these as follows before moving on to discussing the manuscript in Ref. [4].

Remaining subexponential — Tensor networks are applied to efficiently (subexponentially) approximate expansion coefficients c_j . To remain subexponential we must therefore prescribe computations in the space of tensor networks without ever forming the actual coefficients in the usual sense. Additionally, every such computation (as those described below) will tend to increase the auxiliary bond dimensions of the tensors which increases the resource consumption. To keep these in check we must therefore apply truncation schemes (SVD, QR-decomposition, or similar) and fortunately this is an feasible prospect. In many computational instances there is also an optimal order of operations, e.g. in matrix contractions, which can make or break the computationally feasibility. One inescapable caveat, however, is that high entanglement states, e.g. obtained at long durations, can fundamentally not be accurately portrayed (subexponentially) since their continual truncation yields unphysical results.

Non-uniqueness and gauge freedom — An important insight is that the MPS representation of a state is not unique which is evident by our discovery of three distinct canonical forms. Alternatively, it can be seen by inserting $X^{-1}X = I$ for any unitary X into the chain of matrix products. This *gauge degree of freedom* can be used to make nearly all computations easier or even trivial. In particular, it is used prodigiously to gauge mixed-canonical MPS, i.e. shift the *central site*, l , in Eq. (7.21) such that all matrices to the left are left-normalized and all matrices to the right are right-normalized,

$$I = \sum A^{i\dagger} A^i \quad \text{for all } i < l, \quad (7.25a)$$

$$I = \sum B^j B^{j\dagger} \quad \text{for all } i > l. \quad (7.25b)$$

If this was not possible, MPS would not be a feasible state representation and the importance of the non-uniqueness theorem cannot be overstated.

Operators, expectation values, ... — A tensor network representation of a state is by itself of little value without associated standard computations. For example, we can obtain matrix product operator (MPO) representations with the same decompositional ideas

$$\hat{O} = \sum_{j,j'} \langle j | \hat{O} | j' \rangle | j \rangle \langle j' | = \sum_{j,j'} \hat{O}_{j,j'} | j \rangle \langle j' |, \quad (7.26a)$$

by thinking the $\hat{O}_{j,j'}$ as a rank $2L$ tensor decomposable into L rank 4 tensors,

$$\hat{O}_{j,j'} = W^{j_1,j'_1} W^{j_2,j'_2} \dots W^{j_{L-1},j'_{L-1}} W^{j_L,j'_L}. \quad (7.26b)$$

There are other potentially more useful ways to represent operators and in similar regards there are also important considerations of “good” quantum numbers and symmetries [275].

Another standard computation is the state overlap,

$$|\psi\rangle = \sum_j \psi_j |j\rangle = \sum_j M^{j_1} M^{j_2} \dots M^{j_L} |j\rangle, \quad (7.26c)$$

$$|\phi\rangle = \sum_j \phi_j |j\rangle = \sum_j \tilde{M}^{j_1} \tilde{M}^{j_2} \dots \tilde{M}^{j_L} |j\rangle, \quad (7.26d)$$

$$\Rightarrow \langle \phi | \psi \rangle = \sum_{j_L} \tilde{M}^{j_L \dagger} \left(\dots \left(\sum_{j_2} \tilde{M}^{j_2 \dagger} \left(\sum_{j_1} \tilde{M}^{j_1 \dagger} M^{j_1} \right) M^{j_2} \right) \dots \right) M^{j_L}. \quad (7.26e)$$

These nested sums seemingly incur a large computational overhead. If M^{j_i} is left or right normalized, however, then, e.g., we immediately obtain $\langle \psi | \psi \rangle = 1$ and if the MPS is in mixed-canonical form centered on M^{j_i} it is easy to show $\langle \psi | \psi \rangle = \text{tr} \left(\sum_{j_i} M^{j_i \dagger} M^{j_i} \right)$ which entails only a summation of a single set of site tensors because the rest sum to the identity. This exemplifies the strength of gauge freedom and all computational algorithms typically maintains carefully the central site and overall normalization of the MPS.

Diagonalizing — In a somewhat chicken-or-the-egg situation, we do not have c_j for some desired state at our disposal, yet we must somehow find an appropriate MPS representation for it. How can we achieve this? The physical indices j_l and thus the number of site tensors $M^{j_i}_{a_i, a_{i+1}}$, equivalently matrix collections M^{j_i} , are dictated by the physical system under consideration. However, the auxiliary bond dimensions a_i, a_{i+1} (the size of the matrices) and the stored matrix elements are malleable due to the non-uniqueness theorem. Thus, in resolving the conundrum we needn't look further than one of the main topics of this thesis, i.e. local iterative nonlinear optimization where the MPS parameters are considered variational. As always we need to decide on an appropriate cost functional that captures the features of interest. A typical example lies in

determining low lying eigenstates i.e. diagonalization. Proposing a structure and initial content for the matrices (what we would call a seed in dynamical control), we can for example variationally minimize the energy or variance

$$\min_{\{M^i\}_{i=1}^L} E(\psi), \quad \min_{\{M^i\}_{i=1}^L} \text{var}(\psi),$$

respectively, of the MPS by derivative-based means. Gradients, such as $\nabla_{M^i} \text{var}(\psi)$, can e.g. be determined by Gâteaux variational means cf. Sec. 4.3. Often the seed is completely randomized and, because the set of all parameters is a huge search space, the optimization is taken with respect to the individual matrices in sequence (shifting the central site along as described above).

Updating all matrices once is called a sweep and we continually sweep back and forth until converging to a local minimum. As always there is no guarantee that this is also a global minimum and typically we must try out different seeds and perhaps bond dimensions until the best results do not change notably. We then take this as the final MPS representation which can subsequently be suitably gauged for following computations. This procedure works remarkably well despite only optimizing a single matrix (extension to two or several is conceptually straightforward) at a time and this attests to the powerful expressiveness and non-uniqueness of the MPS form. Sometimes this algorithm is called DMRG for historical reasons although we quite clearly did not require new ideas therefrom for this exposition.

Time evolution — Describing the dynamical time evolution is as always central to optimal control and there are several standard algorithms to this end. Inconveniently, a subset of these are really the same algorithm [336] under the guise of increasingly puzzling names (t-MPS, t-DMRG, time-evolution with block decimation, ...). We will go somewhat into detail with one of these in the next section and this turns out to fit perfectly with our Trotterization schemes in Sec. 4.4 yielding efficient exact derivatives. The central idea is that it is difficult/expensive to apply operations across the *entire* MPS but easy/cheap to apply operations on a local subset of site tensors when utilizing proper gauging techniques. There are also other fundamentally distinct algorithms (in more than name) based on Krylov spaces and the so-called Time-Dependent Variational Principle (TDVP) [337].

Boundary conditions — In this presentation we have assumed boundary conditions where either end of the finite chain are uncoupled. There exist parallel MPS developments [275] for which the chain is closed into a ring (periodic boundaries allow treatments of infinite systems to some extent).

Notation — On a less critical note, tensor indices are as mentioned earlier unique identifiers: A^{j_1} belonging to site 1, A^{j_2} belonging to site 2, etc., are *different* sets of matrices. It would be more precise to also attach a specific label for the sites: $A^{[1]j_1}, A^{[2]j_2}, \dots, A^{[L]j_L}$. This is more explicit, but also more cumbersome notation.

Additionally, there are also different approaches to reaching equivalent MPS descriptions, e.g. by expanding the arbitrary state more directly in terms of the Schmidt decomposition [324].

7.4 MPS TIME EVOLUTION AND EFFICIENT EXACT DERIVATIVES

Here we briefly derive a time evolution algorithm that is relevant for the following manuscript. Recall from Ch. 4 that exact derivatives are most easily calculated for

$$\hat{\mathcal{U}}_n^{\text{ST}} = \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2}, \quad (7.27)$$

$$\hat{\mathcal{U}}_n^{c/2} \equiv e^{-i\hat{H}_n^c \delta t/2}, \quad \hat{\mathcal{U}}^d \equiv e^{-i\hat{H}_n^d \delta t}, \quad (7.28)$$

where $\text{ST} = \text{ST}_1$ and $\hbar = 1$. The drift $\hat{\mathcal{U}}^d$ is assumed to include only nearest-neighbor interaction. A time evolution similar to that of Ref. [338] can expedite this calculation at the cost of an increased dynamical error $\mathcal{O}_f(\delta t^3) \rightarrow \mathcal{O}_f(\delta t^2)$ while the derivatives remain exact. In anticipation of the manuscript Ref. [4] we consider the one-dimensional Bose-Hubbard model of Eq. 2.61b with N_s sites — here nondimensionalized by dividing J_x such that $u_n = U(t_n)/J_x(t_n)$ — such that

$$\hat{\mathcal{U}}_n^{c/2} = \exp \left(-i \left(\frac{u_n}{2} \sum_{i=1}^{N_s} \hat{h}_{[i]}^U \right) \frac{\delta t}{2} \right) = \prod_i^{N_s} \hat{\mathcal{U}}_{n,[i]}^U, \quad (7.29a)$$

where $\hat{\mathcal{U}}_{n,[i]}^U = \exp(-iu_n \hat{h}_{[i]}^U \delta t/4)$ is a so-called one-site gate. In this notation, (un)bracketed indices denote (time) sites. This is an exact decomposition because all on-site terms $\hat{h}_{[i]}^U$ commute. The same cannot be achieved for the drift Hamiltonian, however, and we instead obtain

$$\begin{aligned} \hat{\mathcal{U}}^d &= \exp \left(-i(\hat{H}_{\text{even}}^d + \hat{H}_{\text{odd}}^d) \delta t \right) \approx \exp \left(-i\hat{H}_{\text{even}}^d \delta t \right) \exp \left(-i\hat{H}_{\text{odd}}^d \delta t \right) \\ &= \left(\prod_{i \text{ even}}^{N_s-1} \hat{\mathcal{U}}_{[i,i+1]}^{J_x} \right) \left(\prod_{i \text{ odd}}^{N_s-1} \hat{\mathcal{U}}_{[i,i+1]}^{J_x} \right), \end{aligned} \quad (7.29b)$$

with error $\mathcal{O}(\delta t^2)$ where $\hat{\mathcal{U}}_{[i,i+1]}^{J_x} = \exp(-i\hat{h}_{[i,i+1]}^{J_x} \delta t)$ is a two-site gate. The trick here [275, 337] was to first group the sites into even/odd terms (with respect to their lower left index),

$$\hat{H}^d = \hat{H}_{\text{even}}^d + \hat{H}_{\text{odd}}^d = \sum_{i \text{ even}}^{N_s-1} \hat{h}_{[i,i+1]}^{J_x} + \sum_{i \text{ odd}}^{N_s-1} \hat{h}_{[i,i+1]}^{J_x}. \quad (7.30)$$

The first Trotter splitting of these terms induce an error since $[\hat{H}_{\text{even}}^d, \hat{H}_{\text{odd}}^d] \neq 0$, but the second splitting does not due to total internal commutativity of elements in $\hat{H}_{\text{even(odd)}}^d$. Now we combine Eqs. (7.29a)-(7.29b) and shuffle the odd (even) two-site gates to the left (right) until they encounter a non-commutative single-site gate (i.e. they share a site index). That is, for even⁹ N_s , we obtain an approximate $\mathcal{O}(\delta t^2)$ decomposition,

$$\begin{aligned} \hat{\mathcal{U}}_n^{\text{ST}} &= \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}_n^d \hat{\mathcal{U}}_n^{c/2} \approx \prod_i^{N_s} \hat{\mathcal{U}}_{n+1,[i]}^U \left(\prod_{i \text{ even}}^{N_s-1} \hat{\mathcal{U}}_{[i,i+1]}^{J_x} \prod_{i \text{ odd}}^{N_s-1} \hat{\mathcal{U}}_{[i,i+1]}^{J_x} \right) \prod_i^{N_s} \hat{\mathcal{U}}_{n,[i]}^U \\ &= \hat{\mathcal{U}}_{n+1,[1]}^U \left(\prod_{N_s-1}^{i \text{ even}} \hat{\mathcal{U}}_{n+1,[i]}^U \hat{\mathcal{U}}_{n+1,[i+1]}^U \hat{\mathcal{U}}_{[i,i+1]}^{J_x} \right) \hat{\mathcal{U}}_{n+1,[N_s]}^U \left(\prod_{i \text{ odd}}^{N_s-1} \hat{\mathcal{U}}_{[i,i+1]}^{J_x} \hat{\mathcal{U}}_{n,[i+1]}^U \hat{\mathcal{U}}_{n,[i]}^U \right) \\ &\equiv \hat{\mathcal{U}}_{n+1,[1]}^U \left(\prod_{N_s-1}^{i \text{ even}} \hat{\mathcal{U}}_{n+1,[i,i+1]}^{UUJ_x} \right) \hat{\mathcal{U}}_{n+1,[N_s]}^U \left(\prod_{i \text{ odd}}^{N_s-1} \hat{\mathcal{U}}_{n,[i,i+1]}^{J_xUU} \right) \quad (7.31) \end{aligned}$$

$$\equiv \hat{\mathcal{U}}_{n+1,[1]}^U \hat{\mathcal{U}}_{\text{backsweep}} \hat{\mathcal{U}}_{n+1,[N_s]}^U \hat{\mathcal{U}}_{\text{frontsweep}}. \quad (7.32)$$

Here we further grouped triples of operators, e.g. $\hat{\mathcal{U}}_{n,[i,i+1]}^{J_xUU}$, that act solely on nearest-neighbor pairs of indices $[i, i+1]$. Now, the one- and two-site gates can be efficiently evaluated and applied (they are diagonal and time independent, respectively, and both are small in dimension) if the MPS is on an appropriate mixed-canonical form described earlier. This is because the operators [275] are more precisely written e.g.

$$\hat{\mathcal{U}}_{n,[i,i+1]}^{J_xUU} = \mathbf{1}_{[1]} \otimes \dots \otimes \mathbf{1}_{[i-1]} \otimes \hat{\mathcal{U}}_{n,[i,i+1]}^{J_xUU} \otimes \mathbf{1}_{[i+2]} \otimes \dots \otimes \mathbf{1}_{[N_s]}.$$

Then, in applying the associated matrix product operator to an MPS with the center site gauged to either i or $i+1$, all site tensor contractions not concerning $[i, i+1]$ trivially yields the identity due to respective normalizations of matrices cf. Eq. (7.25).

The reason for the particular ordering in Eq. (7.32) and the conceptual definitions of $\hat{\mathcal{U}}_{\text{frontsweep}}$ and $\hat{\mathcal{U}}_{\text{backsweep}}$ is that it allows a very efficient strategy for shifting the center site (each shift costs an SVD), first sweeping forward over the odd gates and then backward over the even gates. An example of this is schematized and detailed in Fig. 7.5.

With all these definitions in place we are now ready to examine the details of the manuscript.

⁹ If N_s is odd, replace $\hat{\mathcal{U}}_{n+1,[N_s]}^U \rightarrow \hat{\mathcal{U}}_{n,[N_s]}^U$ in the final expressions.

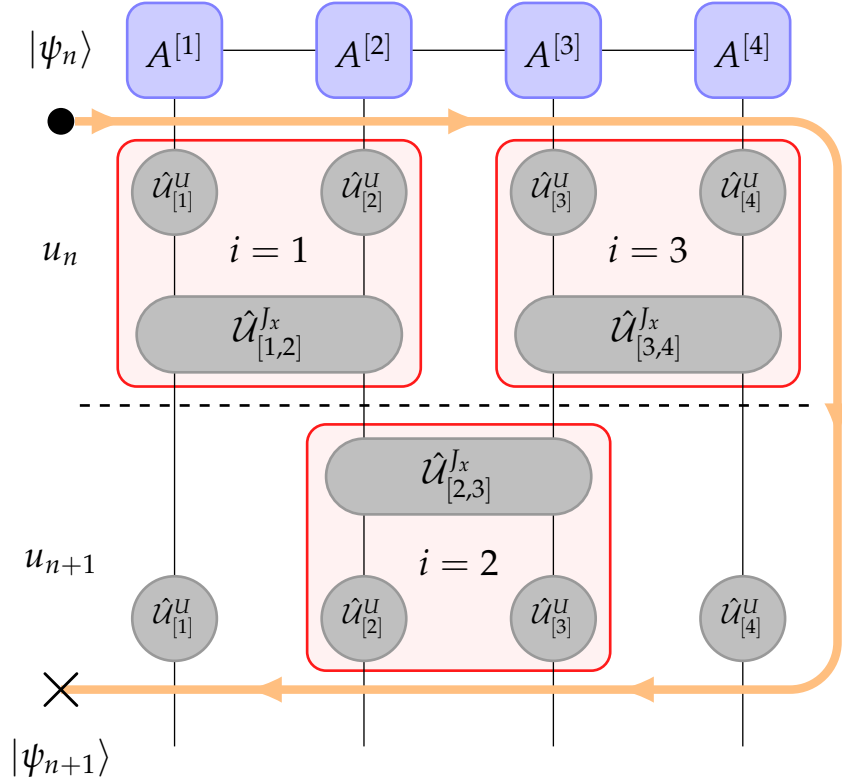


Figure 7.5: Adapted from Ref. [4]. Example diagram ($N_s = 4$) Eq. (7.32) showing the full calculation of $|\psi_{n+1}\rangle = \hat{\mathcal{U}}_n^{\text{ST}} |\psi_n\rangle = \hat{\mathcal{U}}_{n+1}^{c/2} \hat{\mathcal{U}}^d \hat{\mathcal{U}}_n^{c/2} |\psi_n\rangle$ represented as a tensor network diagram [275]. The site tensors (blue nodes) are connected by auxiliary/bond indices (black solid horizontal lines). The one- and two-site gates (gray nodes, time index suppressed) come in triples (red boxes) and are applied to physical indices of the site tensors (black solid vertical lines), in the order indicated by the thick orange-arrowed line (according to the sweeping order). The beginning (end) is marked by a black dot (cross), and the dashed line separates the forward sweep with u_n (above the line) from the backward sweep with u_{n+1} (below the line). Application of product triple i in the forward sweep entails the following:

- (1) Contract the two site tensors with physical indices $i, i+1$ over their common bond index into a temporary two-site tensor,
- (2) Apply the two one-site gates followed by the two-site gate
- (3) Split the temporary two-site tensor by [SVD](#) back into two individual site tensors with the central site (gauge) moved to $i+1$
- (4) Shift the gauge by [SVD](#) an additional site to the right such that the central site is at $i+2$.

Each arrow tip demarcates the gauge position during the sweeps, where sites to the right (left) are right(left)-normalized, with the exception that the next site intersecting the orange line is the central site. After applying the first left-over one-site gate, the backward sweep is similarly performed with the following modifications (1) the site indices are $i-1, i$, (2) the order of gate application reversed, (3) the central site is placed on $i-1$, and (4) the central site is gauged to $i-2$. Finally, the second left-over one-site gate is applied and this completes the time step. Exactly the same procedure with new control values u_{n+1} and u_{n+2} can subsequently be applied to obtain $|\psi_{n+2}\rangle = \hat{\mathcal{U}}_{n+1}^{\text{ST}} |\psi_{n+1}\rangle$. Implementing the backward propagation $|\psi_n\rangle = \hat{\mathcal{U}}_n^{\text{ST}^\dagger} |\psi_{n+1}\rangle$ is similar, but with reversed arrow tips and order of application.

7.5 INTRODUCTION TO THE MANUSCRIPT

In the work of Ref. [4] we apply optimal control techniques (Ch. 4 and Ref. [3]) to drive the superfluid-Mott insulator transition in the one-dimensional Bose-Hubbard model [38, 133, 149, 204, 339–343] in the many-body regime. The states are represented as MPS and time evolution is performed as detailed in the previous section. As described in Sec. 2.5, the superfluid phase is associated with particles tunneling freely while the Mott insulating phase is associated with vanishing particle tunneling and their ground states are, respectively, a site-delocalized superposition and a single¹⁰ Fock state component $|1\rangle^{\otimes N_{\text{sites}}}$ (assuming unit filling).

With the exact diagonalization implementation in Ref. [1] we are feasibly limited to about $N_{\text{part}} = N_{\text{sites}} = 10$ ($D_{\mathcal{H}} \sim 10^5$) which yet requires only a few MB storage memory to store a state (cf. Fig. 7.1) but a lot of computation time. At these system sizes, however, nearly all sites are affected by boundary effects and there are few true “bulk” sites. In this manuscript we consider $N_{\text{part}} = N_{\text{sites}} = 20$ ($D_{\mathcal{H}} \sim 10^{11}$) for which an exact diagonalization state requires about a TB of storage memory. The a MPS representation can be stored in tens of kB, representing a compression factor of roughly $c = 10^8$. Since the transfer is a ground state mapping, each ground state characterized by low entanglement, we expect the MPS to be a good representation (we choose truncation parameters sufficient for describing the intermediate entanglement near the phase transition, see manuscript for the numbers). Generally, MPS have been shown to be efficiently controllable [326] and other ground state mappings have also been investigated [344] with parallels to the paradigmatic two-level LZ model.

It is remarked in passing that the presentation of the manuscript here appears somewhat short compared to the manuscript itself. This is because several discussion topics were promoted to a higher level of detail and treated more independently in the previous sections of this chapter. Additionally, the discussions below are somewhat abridged and leaves out some of the more technical observations.

Before delving into a more quantitative presentation we briefly summarize some of the main results. We optimize the transition to high fidelities $F = 0.99 - 0.9999$ and map out the achievable fidelity as a function of transfer duration T . The optimal control strategies are found to exhibit a bisegmentational structure consisting of (1) an initially somewhat slow ramping across the critical point of the phase transition (population homogenization), followed by (2) a bang-bang

¹⁰ Strictly, there may be a few more components due to finite edge effects but these are heavily suppressed with growing system size.

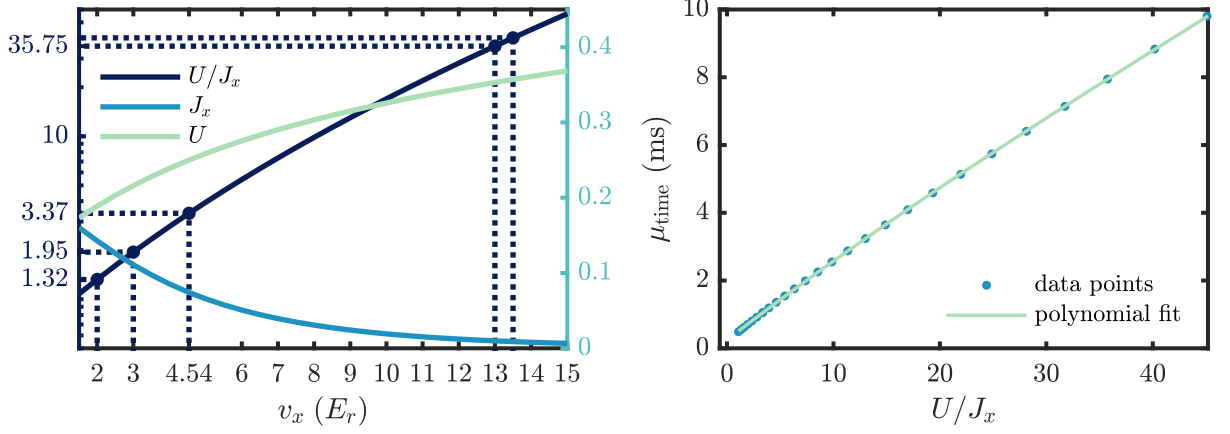


Figure 7.6: Adapted from Ref. [4]. **Left:** Partial reprint of Fig. 2.5. The constitutive relations between v_x and the tunneling J_x and on-site interaction U strengths (multicolored lines and right axis). The ratio U/J_x is also shown (dark blue line and left axis) with a few U/J_x values are highlighted. Left to right they are as per [4]: the minimally allowed $v_x = 2 E_r$, the v_x defining $|\text{SF}\rangle$, the critical point of the one-dimensional $|\text{SF}\rangle \rightarrow |\text{Mott}\rangle$ phase transition, the v_x defining $|\text{Mott}\rangle$, and the “maximally” allowed $v_x = 13.5 E_r$ (corresponding to $U/J_x = 40.18$) **Right:** The conversion factor between SI and simulation units in each discrete time step, $\delta t_{\text{SI}} = \mu_{\text{time}} \cdot \delta t_{\text{sim}}$, depends on $u_n = U/J_x$ and can be accurately fitted by a polynomial.

structure (alternating tunneling and phase adjustment). Each strategy is characterized by an integer number of bangs in the second segment (the first segments are always rather similar) and the best one depends on T . In comparing to similar previous works using derivative-free methods on more lenient optimization objectives than fidelity [73, 74] we find significant quantitative improvements and qualitative differences. More generally, to our knowledge, the contemporary/parallel work by another group resulted in the first published [345] application of gradient-based optimization in MPS using first-order approximate gradients. The work presented here is then possibly the second published gradient-based application in MPS overall, but the first to realize exact derivatives and associated benefits.

7.6 MANUSCRIPT RESULTS

In order to enable application of the Trotterized exact derivatives of Ch. (4) we rescale the original Bose-Hubbard Hamiltonian Eq. (2.61b) with J_x such that

$$\hat{H}_n = \hat{H}^d + \hat{H}_n^c = \sum_{i=1}^{N_s-1} \hat{h}_{[i,i+1]}^{J_x} + \frac{u_n}{2} \sum_{i=1}^{N_s} \hat{h}_{[i]}^U, \quad (7.33)$$

$$\hat{H}^c = \frac{1}{2} \left(\sum_{i=1}^{N_s} \hat{n}_i (\hat{n}_i - 1) \right), \quad (7.34)$$

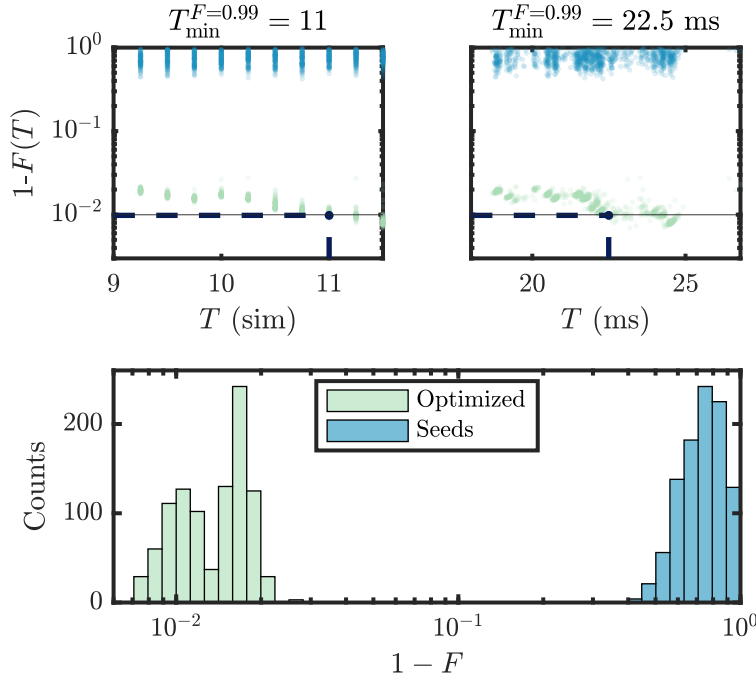


Figure 7.7: Adapted from Ref. [4]. Mapping out the fidelity as a function of duration in the vicinity of $T_{\min}^{F=0.99}$. The top panels show $1-F$ both in simulation and SI units. The bottom panel shows a histogram of all unoptimized and optimized fidelities.

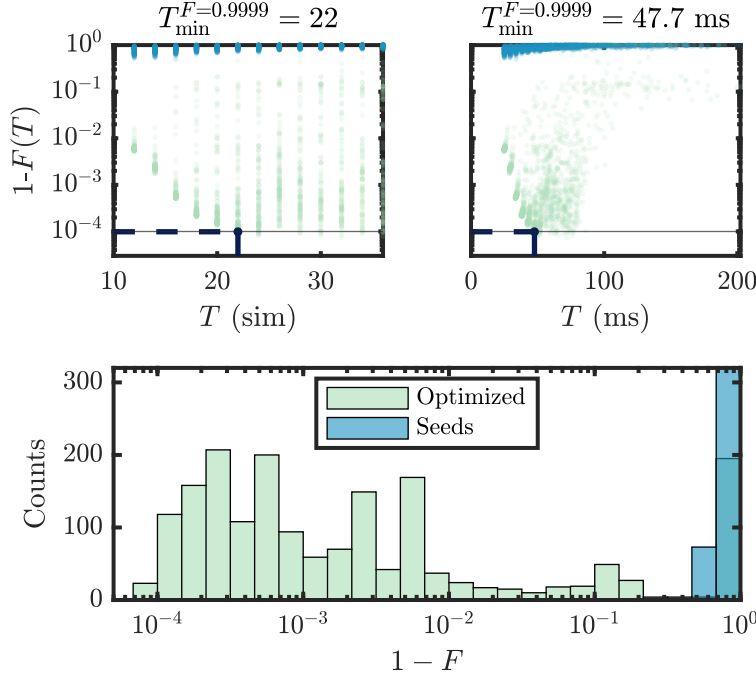


Figure 7.8: Same as Fig. 7.7 but in the vicinity of $T_{\min}^{F=0.9999}$.

where the control parameter is $u_n = U(t_n)/J_x(t_n)$. The drift Hamiltonian $\hat{H}^d$ represents isotropic tunneling between the sites and the control Hamiltonian $\hat{H}_n^c$ represents the on-site interaction energy. The latter and its derivative $\hat{H}^{c'} = \partial \hat{H}_n^c / \partial u_n$ are diagonal in the site occupation basis. The time independence of $\hat{H}^d$ and $\hat{H}^{c'}$ is exploited for computational speed ups using caching techniques.

The simulation and optimization is performed without explicit constitutive relations for U and J_x in mind. This means that the optimal controls $u^* = (U/J_x)^*$ are valid for *any* physical system describable by the model. The only possible encumbrance, if certain comparative studies are desired, is that we must *a posteriori* translate the results into a given physical system by appropriately derived constitutive relations to e.g. allow meaningful durations in SI units. Concretely, the time scale $\mu_{\text{time}} = \hbar/J_x$ between simulation and SI units is variable and depends on J_x such that the full protocol duration in an optical lattice is given by

$$T_{\text{SI}}(\mathbf{u}) = \delta t \sum_{n=1}^{N_t} \mu_{\text{time}}(J_x) = \hbar \delta t \sum_{n=1}^{N_t} J_x^{-1}(v_x(u_n)), \quad (7.35)$$

where N_t is the number of time steps and δt is given in simulation units. See the manuscript appendix for a more detailed discussion. Fortunately we have already calculated these constitutive relations in Sec. 2.5 for an optical lattice, relating $U(v_x)$ and $J_x(v_x)$ to the physically controllable parameter v_x as shown in Fig. 7.6 which also shows $\mu_{\text{time}}(J_x)$.

7.6.1 Control problem

We optimize the state transfer fidelity of

$$|\psi_{\text{ini}}\rangle = |\text{SF}\rangle = |\text{GS}; 3E_r\rangle \rightarrow |\psi_{\text{tgt}}\rangle = |\text{Mott}\rangle = |\text{GS}; 13E_r\rangle, \quad (7.36)$$

where $|\text{GS}; v_x\rangle$ is the ground state at a variable v_x but fixed $v_y = v_z = 20E_r$ depths, see Fig. 7.6 for corresponding U/J_x values. The figure also shows other relevant values for the depth v_x . Some of these are related to modeling assumptions in deriving the model and others are purely chosen for convenience in the optimization or to enable a degree of quantitative comparison to the previous works on similar studies [73, 74]. For the optimizations we use the nonlinear interior-point method implemented in the IPOPT software library [265]. We supply the exact gradients formed by Eq. (4.64a) which are internally used in forming a BFGS Hessian approximation. These and more quantities — e.g. control boundaries — are used in forming the interior-point search direction. Line searching is also handled internally by IPOPT. The control is parametrized as a piecewise constant with full temporal resolution. We implement control regularizations accurate to order

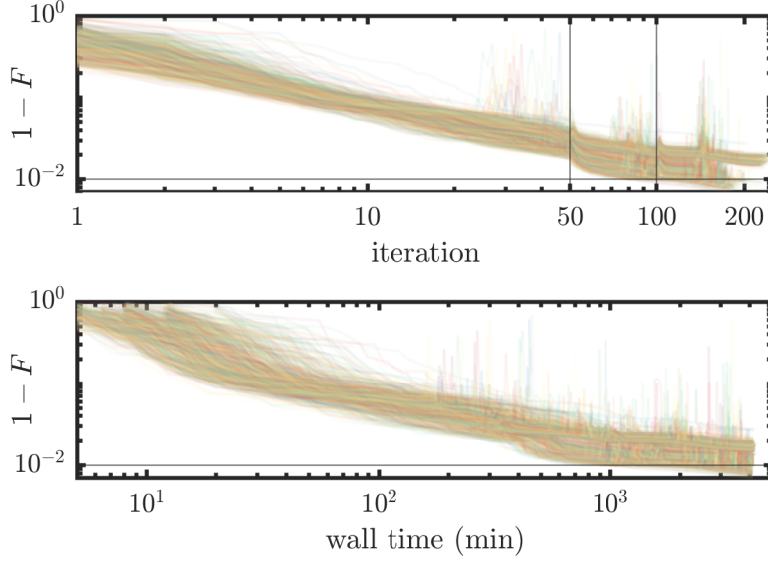


Figure 7.9: Adapted from [4]. Fidelity optimization trajectories in terms of iteration and wall time consumption for the results in Fig. 7.7. The horizontal lines denote $F = 0.99$ and the vertical lines denote the changes in homotopy parameters $\{\delta t\} = \{0.1\} \rightarrow \{0.05\} \rightarrow \{0.025\}$ (cf. Sec. 3.7) which roughly reduces the time scale for optimization from weeks to days. The homotopy sequence is appropriate since all regions follow roughly the same power law. Each iteration on the finest time grid takes on the order of hours.

$\mathcal{O}(\delta t^2)$ and typically small values are used ($\alpha, \gamma \sim 10^{-7} - 10^{-10}$). The tensor network calculations are based on the ITENSOR software library [334] and for the state dynamics we implement the tensor network time evolution algorithm specified in Fig. 7.5.

7.6.2 Optimization results

Figures 7.7–7.8 show the best empirically achieved fidelity as a function of process duration, approximately given by an exponential law, and from these we find minimal duration estimates $T_{\min}^{F=0.99} = 11$ (22.5 ms) and $T_{\min}^{F=0.9999} = 22$ (47.7 ms). The effect of the optimization is seen to be significant as all solutions are improved by several orders of magnitude. The exact gradients are efficiently evaluated and the iteration wall time consumption depends only on the time it takes to propagate the states (i.e. linearly in the $2N_t$ total time steps for the forwards and backwards propagation). Sample optimization trajectories are shown in Fig. 7.9.

Looking at the optimized solutions themselves in Fig. 7.10 we observe two segments with distinct behaviors.

Segment 1 – crossing critical point: $t/T \in [0, \frac{2}{3}]$. Outside an initial perturbation, this segment slowly crosses the critical point with nearly vanishing variance and high-frequency/low-amplitude oscillations.

Segment 2 – bang-bang structure: $t/T \in [\frac{2}{3}, 1]$. With large variance, this segment ramps to the final control value and consists of several low-frequency/high-amplitude oscillations. Upon further study it turns out that these oscillations are, in fact, smoothed approximations (induced by regularization costs) to bangs.

Resolving the optimized solution according to their duration T , focusing on the second segment, and subtracting there the linear contribution in Fig. 7.11 we obtain a clearer picture. Namely, there are several optimal control strategies $C_{N_{\text{peak}}}(T)$ and these are individually characterized by an integer number of peaks/bangs $N_{\text{peak}} = 5, 6, 7$. The best and most prominent one of these, $C^*(T)$ cf. Sec. 3.2.1, depends on T (higher T leads to a higher optimal number of peaks) and, outside of this particular range of investigated durations, different N_{peaks} are also found.

To understand the physical mechanisms at play in the problem we inspect the quantum dynamics induced by fidelity-optimized controls at different durations in Fig. 7.12. Shown there is also the density of defects and the rescaled average variance over all sites N_s ,

$$\rho(t) = \frac{1}{N_s} \sum_{i=1}^{N_s} |\langle \hat{n}_i \rangle_{\psi(t)} - 1|, \quad \eta(t) = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\Delta \hat{n}_i(t)}{\Delta \hat{n}_i(0)}, \quad (7.37)$$

where $\langle \hat{n}_i \rangle_{\psi(t)} = \langle \psi(t) | \hat{n}_i | \psi(t) \rangle$ and $\Delta \hat{n}_i(t) = \langle \hat{n}_i^2 \rangle_{\psi(t)} - \langle \hat{n}_i \rangle_{\psi(t)}^2$. These were used in earlier derivative-free work [73, 74] and these figures of merit are much more lenient than fidelity. For example, many state configurations lead to minimal/low values in addition to the $|\text{Mott}\rangle$ ground state and this effect grows with system size. From $\langle \hat{n}_i \rangle_{\psi(t)}$ in the figure we find that a light-cone-like homogenization of site population takes place when crossing critical point (first segment). The population is otherwise initially concentrated in the bulk (roughly sites 4-17) due to finite edge effects. This process is clearly limited by the propagation velocity $\partial \langle \hat{n}_i \rangle_{\psi(t)} / \partial t$ and thus also influences the minimal duration T_{min}^F for high-fidelity transfers. The tardiness hereof is somewhat expected because we are crossing a phase transition. Briefly, there is the additional complication in the present problem that, in the superfluid limit, the spectral gaps tighten to small finite values $\Delta E \rightarrow 0$ (exactly zero in infinite systems) and this implies very long (diverging) adiabatic time scales $\tau \sim 1/\Delta E$. Thus, residual population/defects may become pinned which can be estimated either by Kibble-Zurek theory or as a cascade of independent LZ avoided crossing transitions [344, 346–348].

Even upon reaching a largely homogenized site population after the first segment, however, the fidelity remains vanishingly small. The

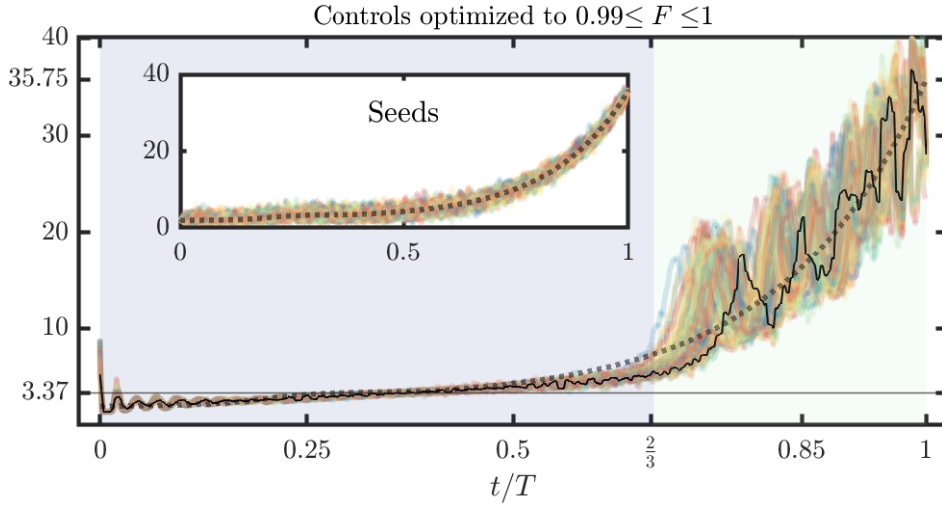


Figure 7.10: Adapted from [4]. High-fidelity optimized controls (translucent multicolored lines) and their seeds (inset). The critical point for the phase transition $(U/J_x)_{\text{crit}} = 3.37$ is marked by the horizontal line. The dotted line is the reference control which is perturbed by trigonometric functions. The bisegmentational structure is highlighted by the blue and mint colored areas.

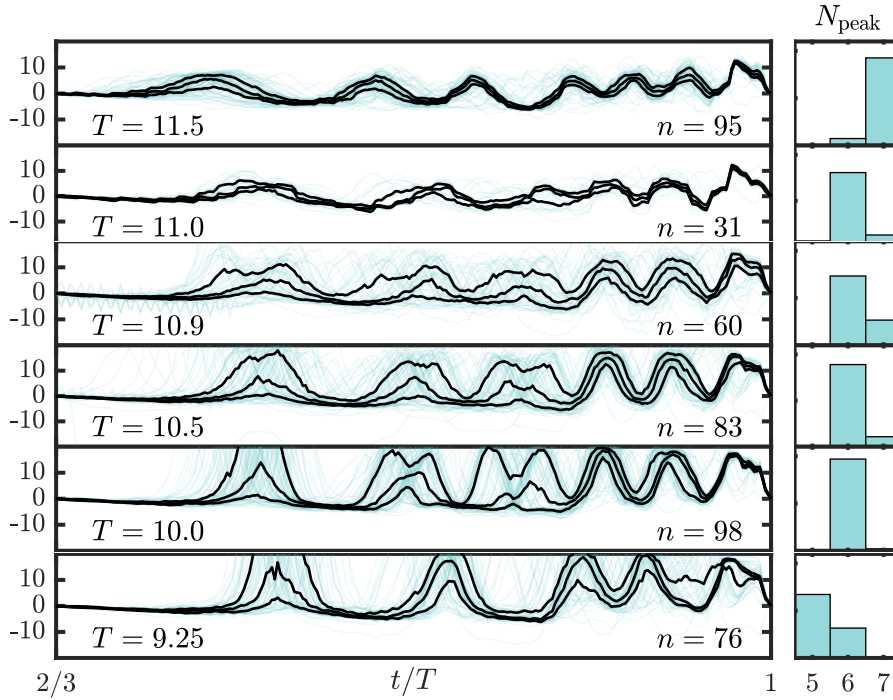


Figure 7.11: Adapted from [4]. Distinct optimal control strategies characterized by a different number of peaks/bangs are active at different T (rows). The left panels show the n found high-fidelity controls (translucent lines) and their (25,50,75)% quantiles (black lines). The right panels show the relative distribution of the n solutions into the three identified strategies.

T (ms)	$1 - F$	ρ	η
20.5	$1.0 \cdot 10^{-2}$	$2.4 \cdot 10^{-4}$	$4.1 \cdot 10^{-3}$
14.9	$0.7 \cdot 10^{-1}$	$2.1 \cdot 10^{-3}$	$1.9 \cdot 10^{-2}$
5.7	$0.4 \cdot 10^0$	$1.9 \cdot 10^{-2}$	$0.9 \cdot 10^{-1}$

Table 7.1: From Ref. [4]. Figure of merit value minima when optimizing fidelity. Note that the ρ, η minima may occur at $t < T$. References [73],[74] report as best achieved results $\rho_{\text{ref}} = 10^{-3}$ and $\eta_{\text{ref}} = 10^{-1}$, respectively optimized for individually with derivative-free means in similar systems.

following bang-bang process (second segment) corresponds to alternating between tunneling events and site-locking — during the latter, individual phases are imprinted on the individual Fock components

$$\hat{\mathcal{U}} |n_1, \dots, n_{N_s}\rangle \approx \exp \left(-\frac{i u^{\max} \delta t}{2} \sum_{l=1}^{N_{\text{sites}}} n_l (n_l - 1) \right) |n_1, \dots, n_{N_{\text{sites}}}\rangle, \quad (7.38)$$

since tunneling is approximately negligible. This leads to a non-trivial interplay (highly non-monotonic $F(t)$) between the canonically conjugate population and phase variables that is apparently required to produce the correct interferences finally leading to the target $|\text{Mott}\rangle$ state. In comparing with Refs. [73, 74] in Tab. 7.1 we find that results therefrom are not quite consistent with high-fidelity transfers (they also do not claim to be). Additionally, when here optimizing for the fidelity, we also find solutions yielding $\eta \approx \eta_{\text{ref}}$ at a third of the duration compared to Ref. [74].

The presented optimized controls also exhibit a higher degree of information-theoretic *control complexity* [326, 349] — that is, a bisegmentational structure with bangs as opposed to a very smooth one¹¹. This statement can be understood both qualitatively and quantitatively.

In conclusion, the exact gradient-based Trotter methodology with diagonal control enabled significant improvements in obtaining and interpreting the optimal control strategies for the paradigmatic Bose-Hubbard phase transition in the many-body limit.

7.7 PERSPECTIVES

A possible critique of the presented results is that $F = 0.99 - 0.9999$ are not currently feasible in experiment or application and are therefore not relevant to consider. In my opinion, however, this type of statement is not particularly robust because it is circumstantially premised on

¹¹ This difference cannot be fully attributed simply to the respective choices of parametrization.

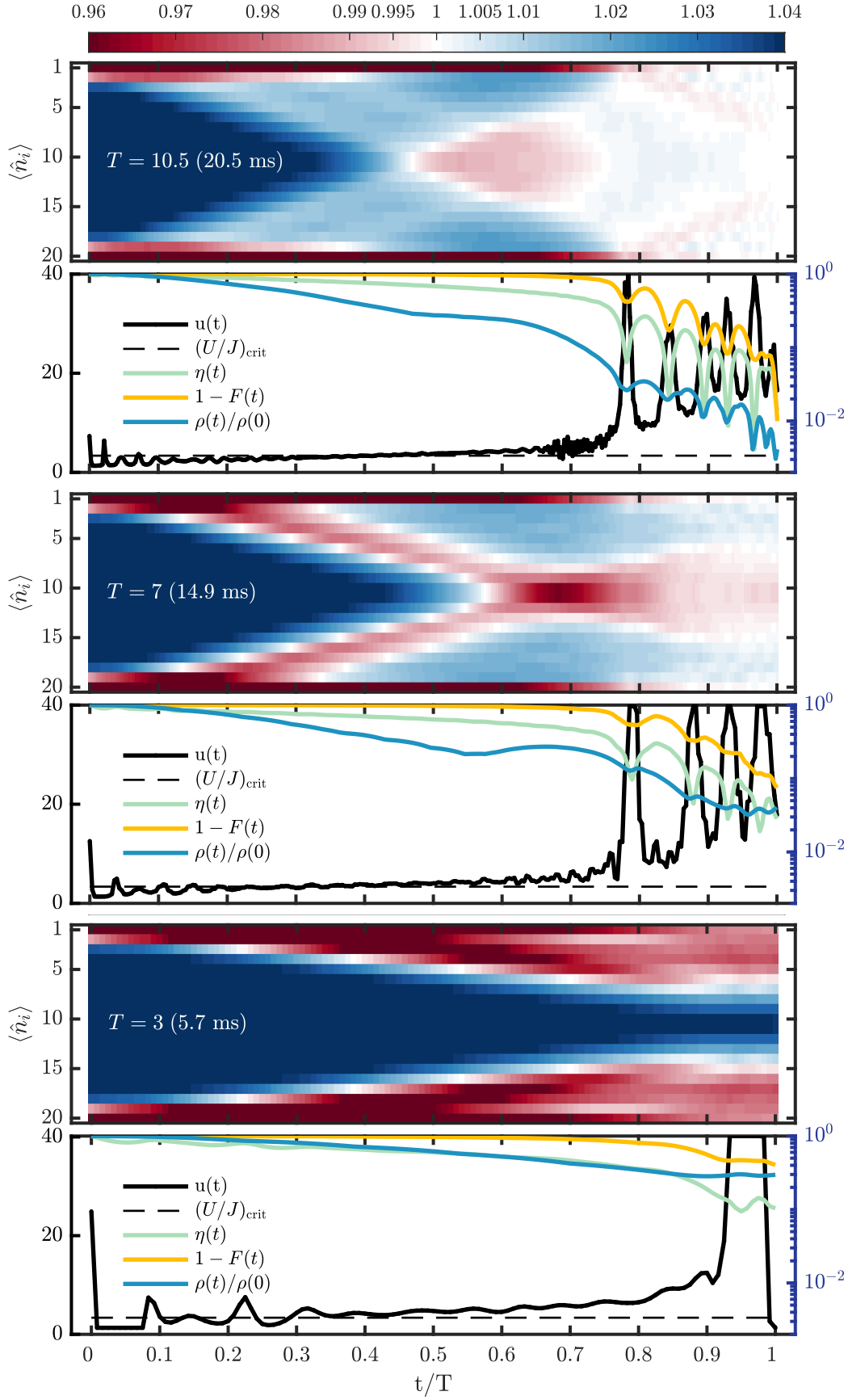


Figure 7.12: Adapted from [4]. Time evolution along optimized controls at three different durations. Each upper panel shows the site occupations $\langle \hat{n}_i \rangle_{\psi(t)}$ (sensitive near-unit occupancy and note $\langle \hat{n}_1 \rangle_{\psi(0)} = 0.67$) while each lower panel shows the corresponding control (black line and left axis) and different figures of merit (multicolored lines and right axis).

experimental limitations and is somewhat dismissive of theoretical advances — and developments in theory and experiment are seldomly in perfect lockstep. In the spirit of the introductory 1952 quote by Schrödinger in Ch. 1, it is impossible to say if and what experimental limitations will continue to be relevant and for how long. Regardless, in the meantime, open-loop methodologies may actively be a source of worthwhile contributions as discussed to a somewhat lengthy extent in Sec. 3.8. In the optimized superfluid-Mott transition, for example, the presented identification of bisegmentational optimal control strategies provides an obvious parametrization for a closed-loop optimization in a significantly truncated search space that can empirically close in on the high-fidelity regime $F \geq 0.99$ (granted that the control quality actually deteriorated in experiment in the first place).

On another note, it would be conceptually trivial and also numerically feasible to include robustness through ensemble optimization. This could be in terms of run-to-run fluctuations in total atom number or any other characteristic error source.

Let us close by discussing certain future avenues for application of the presented control methodologies. Since the exact derivatives are available for the same price as evaluating the system dynamics, it is the latter's state-of-the-art that now primarily sets the limits of feasibility for control. For example, it would be very interesting to consider quantum many-body systems in more than one dimension. Following the associated entanglement area laws, however, the MPS representation is typically no longer well-suited and we must instead employ alternative formulations. Currently, there seems to be two prominent research directions in this regard.

The first direction consists in increasingly advanced tensor networks such as projected entangled pair states (PEPS), tree tensor networks (TTN), or the multi-scale entanglement renormalization ansatz (MERA) [56, 328, 332, 350–352]. Some of these are also able to treat nonzero thermal components [353].

The second direction follows the machine learning trends of the past decade. Namely, neural networks are another type of variational state ansätze that have recently garnered attention [354–361]. A neural network nn is simply an expressive mathematical object (amenable to a pictorial network depiction) parametrized by internal weights W that can be optimized (or “trained”) to approximate some function, $nn(x; W) \approx f(x)$, over appropriate ranges of input x . Typically $f(x)$ is expensive or otherwise unfeasible to evaluate while $nn(x)$ is not. The nn expressiveness is given by the (tunable) number of weights W and are usually universal approximators (under reasonable assumptions, see e.g. Ref. [362] and similar works for other architectures). The idea, then, is that a properly trained neural network, in strong resemblance to tensor networks, is an efficiently compressed representation of

the original function which could be the complex state coefficients $c_j = c(\mathbf{j})$ and their time evolution. To my knowledge, optimal control of such approximation schemes have not yet been presented in the literature and these may also be able to extend into the higher-dimensional quantum many-body arena. Perhaps unsurprisingly just from the very common terminology and analogous concepts, neural networks can be mapped into tensor networks [363].

On a somewhat unrelated note, machine learning techniques have also been applied in mapping out phase diagrams [359] and for quantum state tomography [364] (reconstructing potentially complex physical quantities from relatively simple measurements).

MANUSCRIPT [4] – ACHIEVING FAST HIGH-FIDELITY CONTROL OF MANY-BODY DYNAMICS

Manuscript in preparation.

16 pages.

AUTHOR CONTRIBUTIONS

This project was initially worked on by Frederik Møller (during his master's project) and Jens Jakob Sørensen. These performed initial numerical studies and achieved proof of concept results in a C++ code. At that time, there was no delineation between this project and what has since evolved into manuscript [3].

When I continued the project, I rewrote large portions of the existing C++ code and added more robust unit testing. Among others, the forwards and backwards time evolution were improperly implemented (they are fairly involved as shown Fig. 8 of the manuscript) which yielded inexact gradients. I also expanded the scope with several new features. For instance, I devised mechanisms for saving and loading states and operators — which are very expensive calculations in the many-body limit — at runtime. I also implemented flexible homotopy (sequential handover between different levels of optimization coarseness) techniques through certain serialized input structures. These mechanisms reduced, roughly, the time scale for individual optimization runs from weeks to days, making the necessary large scale studies much more feasible. Additionally, I calculated (see manuscript [3]) and implemented new exact regularization gradients.

I performed the numerical studies for the paper, and conversion analysis between simulation and SI units, and analyzed and interpreted the results.

It became clear that the significantly increased scope warranted two separate publications, this and the manuscript of Ref. [3].

In the writing phase, I authored the manuscript and figures with scientific feedback and corrections from Frederik Møller and Jacob Sherson.

WHAT it means to *solve* a quantum optimal control problem can, broadly speaking, be understood in two ways as described in Ch. 3: either as finding a single (or few) solutions realizing $F \approx 1$ or as systematically characterizing the underlying optimal control strategy(ies) $C^*(T)$. Correct identification of the latter leads to a principally complete description of the problem but is also an order of magnitude or more challenging and time consuming to achieve. The reason for this inherent additional difficulty lies partially in that we need to devise seeding strategies, S , that are both focused (exploitative) enough to densely probe promising regions of the control landscape — where we might *suspect* $C^*(T)$ is — but also broad (exploratory) enough to not just repeatedly sample “the same points”. That is, the S must sufficiently contain $C^*(T)$ as depicted in Fig. 3.9.

8.1 CONTROL PROBLEM COMPLEXITY

The stone in the shoe is that we cannot, for sure, know exactly how to choose and balance these seeding aspects, and outside of elusive analytical insights we only have license to speak and conclude on empirical results attained through our optimizational efforts. Nevertheless, if this stone is a pebble or a boulder depends on the problem complexity which we can in broad terms divide into two separate axes: the numerical complexity and the inherent topographical complexity of the control landscape.

Landscape complexity is a property of the control problem itself due to the cost functional mapping. It relates to the landscape topography, e.g., the multiplicity of local traps, global optima (optimal strategies), and their locations, sizes, correlations, as well as other “geometric” features cf. Sec. 4.6. Qualitatively, if this complexity is low then the optimal strategies are broad attractors and the local traps few and far between and conversely if the complexity is high.

Numerical complexity is due to the amount of “compute” ($\sim$ the computation time) required by control algorithms to faithfully traverse the landscape in seeking troughs or peaks. We have seen in Ch. 4 that certain algorithmic choices reduce the complexity for exact derivatives, yet we must still solve the equations of motion many times over even in this case. This complexity therefore naturally scales with the system size under consideration.

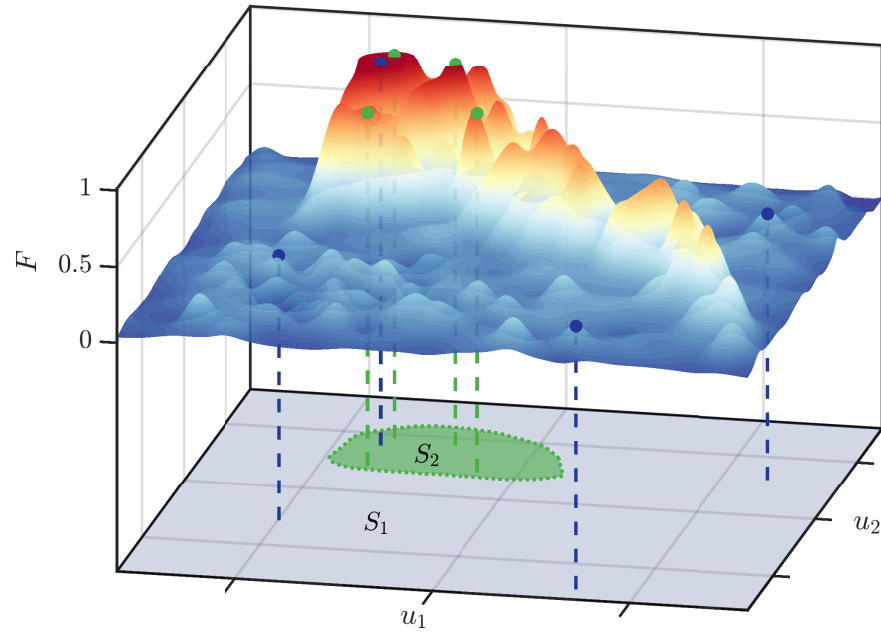


Figure 8.1: Re-print of Fig. 3.9 showing a control landscape with a single optimal control strategy covered partially by two different seeding strategies.

The *overall* problem complexity is then, in this admittedly simple view, a sort of product of the two and this leads to different possible scenarios as follows.

If the system is sufficiently small and the equations of motion “simple”, thus requiring minimal compute per optimization iteration e.g. order of milliseconds as in the examples of Ch. 4, we may conceivably perform a very wide, dense, and simple uniform random probing of the landscape (S_1 in Fig. 8.1) to find $C^*(T)$ within a reasonable time frame. In this limit the topographical complexity is (approximately) inconsequential and decent to great results may be produced even by “bad” seeding and algorithmic approaches.

As the numerical complexity increases to the opposite limit, however, this naïve “wide cast net” approach likely becomes much less feasible since progress with every iteration (each e.g. on orders of hours as in Ch. 7 or more) becomes correspondingly more important. Yet, if the landscape complexity is sufficiently simple, e.g. containing a single broad attractor with no local traps whatsoever, the achieved results will not suffer significantly beyond some “minimal effort” threshold.

However, if both numerical *and* landscape (and thus overall) complexity becomes significant it is much preferable or necessary to construct specialized seeding strategies (S_2 in Fig. 8.1) drawing typically on the accumulated knowledge, expertise, and programming proficiency of the domain expert. This can, however, be a challenging prospect for the reasons stated in the introductory paragraph and we

can then ask the question: are there possibly any viable alternatives to the conventional methodologies? Would it, for example, be possible to locate the “good” region S_2 by *crowdsourcing* the quantum control problem to nonexpert citizen scientists through an appropriate interface? And, further, how does such potential prospects depend on the problem complexity? That is the topic of this chapter.

8.2 CITIZEN SCIENCE

Citizen science [365, 366] methodologies — i.e. enlisting nonexpert citizens from the population at large to help solve scientific problems — have in recent times been successfully applied in the natural sciences and adjacents. This includes, e.g., “offline” activities [365] in ecology like bird counting, studies of the evolution of species, and various environmental and geographic science [367], as well as “online” activities such as general image labeling [368], quantum Bell tests (The BIG Bell Quest) [369], the creation of BEC [81] (Alice), classification of galaxies (GalaxyZoo) [370], DNA sequence alignment (Phylo) [371, 372], 3D reconstruction of retinal neurons (EyeWire) [373], building new RNA structures (EteRNA)[374], and protein folding problems [375] (Foldit) [376, 377]. Although this is a quite diverse list of applications, they are all by and large enabled by the rapid communication channels of today’s day and age which allows researchers to efficiently crowdsource [378, 379] certain elements of a given scientific study. We shall focus henceforth on the “online” activities.

Citizen scientists can add scientific value in several different ways. Perhaps most prominently is the express utilization of human cognitive functions in solving computationally hard problems. For example, human pattern-recognition skills enable quick solutions to certain classification and processing tasks [368, 370] that to us may be relatively “simple” although they are in fact highly nontrivial from mathematical point of view. Other potential applications draw on higher order functions, such as creative [376], new, abstract, critical, or emotional thinking. For some of these, e.g. in so-called *complex problem solving* [380, 381], even stating the problem in a mathematically well-defined way is challenging or impossible. Any discussion and celebration of human qualities, of course, extends far beyond the limited scope of citizen science.

On the other hand, with increasingly powerful computational resources and, perhaps of especial note, the rise of machine learning algorithmic paradigms in the past two decades we have also experienced that human capacities, no matter how romantic, can be rivaled in many areas. Highly publicized examples are Deep Blue [382] and [383] AlphaGo [384], defeating human world champions in the games of chess and go, respectively, and the latter was shortly after even defeated by a *tabula rasa* version of itself, AlphaGoZero, just a year

later [385]. These and other similar success examples tend to have rigorous and “simple” rules from which the problem complexity may be emergent. Perhaps more closely related to serious real-life concerns are *expressly* simple tasks with clear algorithmic solutions. For example, tasks of monotonous or tedious nature carried out by humans are prone to human error [386] whereas the same task carried out by a computer will “never” fail.

Nevertheless, likewise romanticisation of machine learning and AI approaches may be questioned [387, 388] for reasons related to the preceding paragraph and others. Of particular note is the lack of “common sense” comprehension which is clearly a major roadblock for general-domain AI [389]. In the manuscript [5] we define this somewhat elusive quality as the “*implicit and shared fundamental assumptions that people have about the world* [390]”. From a certain point of view, common sense is what allows humans to apply the same logic, i.e. transfer our knowledge [389], between similarly natured tasks.

Having now argued the strengths and weaknesses of both human and computer algorithmic capacities¹ it seems plausible that integrated

¹ Consider solving the following task:

A very thin glass containing 1 deciliter of water sits on a 1.5 meter tall, old, cluttered table with four legs which stands on a concrete floor. A cat leaps out of a nearby box, containing an assortment of radioactive materials and poisons, onto the table and hits the glass. The glass slides over the edge of the table. What happens next?

Answer: the glass falls to the ground and shatters

The solution to this problem is trivial for an average human and is immediately identified after reading the description. Although a lot of information is present, it is obvious to us that only few pieces are relevant (the glass is no longer in contact with the table, thin glass is brittle, concrete is solid). What is more, we have also inferred reasonable basic assumptions about the situation not *explicitly* present in the formulation (the object relevant for the question is the glass, it takes place on Earth so the glass will fall to the ground, ...) and the question itself is rather open-ended. It is also obvious how the solution would change, e.g., if the table is 1cm tall and the floor is made of rubber — or would not change, e.g., the cat leisurely strolls by and paws the glass of the table or if the text color is altered [391]. That is, we are able to apply the same logic, or transfer our knowledge, between similar problem tasks because of our “common sense” that we have acquired through experiencing the world. On the other hand, these solutions and qualities would be much harder to capture by a computer algorithmic approach, see e.g. the even simpler examples in [389, 391].

Conversely, there are also many tasks in which humans will with all likelihood fail [392]. Consider, for example, the following computationally simple problem:

Remember this number:

3129812378621342341234140741234341386109

Answer: 3129812378621342341234140741234341386109

This is a task that is trivial for a computer, but hard for the average human. The computer will always outperform in this and similar problems because they are primarily designed exactly for solving them using a set of mathematical and/or algorithmic rules or instructions. Note that these two examples are merely illustrative.

synergistic systems, in so-called *hybrid intelligence* [393–395], could yield significant advantages. Serving perhaps as the most simple, yet concrete, example is to inject human input [396] into one or several parts of an otherwise well-defined algorithmic loop and this is often aptly referred to as human-in-the-loop [397–401]. Returning once more to the explicit citizen science context, this is the predominantly realized model because the sought after human input is not necessarily related to domain specific knowledge but, rather, the overall cognitive apparatus and/or “common sense” comprehension possessed by the population at large.

The idea is then to carve out one or more, possibly very simple or very complex, spaces in the scientific process and provide citizen scientists with a limited and suitable interface hereto. The “online” activities mentioned in the introductory paragraph of this section are examples of such interfaces. In general, it is in itself a challenge [402–404] to identify and implement the optimal set of interface features and functionality that best mediates both (1) the solution of the problem itself [interest of the scientists], and (2) instilling enjoyment and/or a feeling of making valuable contributions [interest of the users]. Both of these aspects are important: the first for obvious reasons and the second because citizen scientists are typically enthusiasts that volunteer their time with the implicit expectation of some extrinsic or intrinsic reward. Related to the second point, many interfaces are framed as games to make the experience more palatable and in line with a subset of certain user expectations. On a somewhat tangential note, citizen science projects are also great outreach opportunities for increasing scientific literacy and awareness [366].

As a possible arena for crowdsourcing, let us narrow in once more on the quantum optimal control task in Fig. 8.1. We can now ask several related questions in more concrete detail than before: is it possible solve this problem using hybrid intelligence, in the human-in-the-loop sense, where citizen scientists formally take the role of adaptive seeding strategies? And can they, in cooperation with concrete standard optimization algorithms, locate the “good” region(s) S_2 of input space? That is, is it feasible that introducing and crowdsourcing “human common sense” in the algorithmic loop can yield potential benefits over or complementarity to baseline methodologies?

8.3 INTRODUCTION TO THE MANUSCRIPT

As a means to study possible alternatives to the “conventional”, fully-expert driven solution of optimal control problems, the manuscript [5] reports on our citizen science game QUANTUM MOVES 2 described in Fig 8.2. There, nonexpert citizen scientists, henceforth referred to as players, can produce seeds in real-time and subsequently engage with a local optimizer embedded in the game that runs fully on their device.

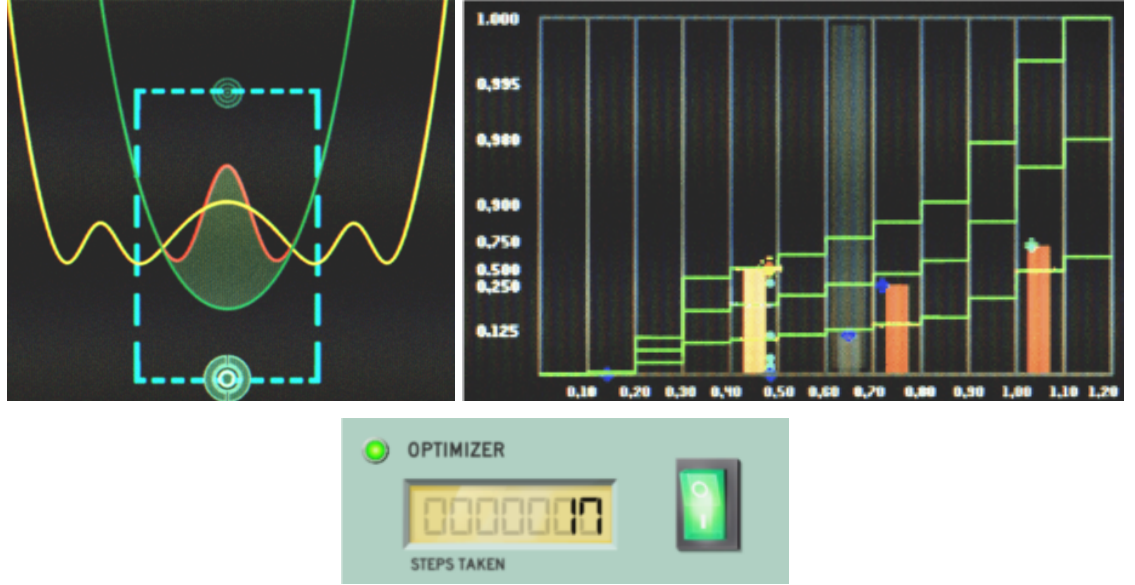


Figure 8.2: Adapted from Ref. [5]. Partial screenshots of the QUANTUM MOVES 2 game interface. Top left: The play view (here showing the *Splitting* problem) where the player can produce seeds. The mouse cursor coordinates (solid turquoise circle) is mapped to relevant control parameters and changes, upon being moved, the instantaneous potential (green line). The wave function (red line w. shaded area) is propagated in real-time and the goal is to reach the target state (yellow line). The states are displayed on top of the potential for visual purposes only. The control functions can have any shape but are clamped to within the bounding area (turquoise dashed box). Top right: The $F(T)$ view. The x axis is divided into 12 bins normalized by a suitable reference duration near $T_{\min}^{F=0.99}$. The y axis shows (nonlinearly) the fidelity. The player can inspect and select any of their produced solutions (dots) for optimization. The upper green line was produced by conventional methods prior to launching the game and the lower green two lines are motivational gameplay elements (score is aggregated based on relative placement to these). Bottom: The player can start and stop the local optimizer by clicking the button and the views are updated with each iteration. Additional features such as solution replay and side-by-side *ghosts* are not shown here.

We gather the seeds and optimized solutions thereof and compare their large scale bulk efficacy to optimizations performed with piecewise constant uniform random seeding. Uniform random seeding is the most heuristic free, generalistic, and, by natural corollary, also the most naïve form of seeding which may or may not be sufficient depending on the overall problem complexity as discussed in Sec. 8.1. The results are therefore to be understood as a baseline for feasibility. It is not meant to promote an inimical relationship or competition between humans and algorithms, but rather entice their cooperation in forming the so-called hybrid intelligence [393, 394] — drawing on the best aspects of both in complementarity.

More concretely and introducing some notation, we solve three optimal control problems — the single-particle and two meanfield

BEC problems schematized in Fig. 8.3 — using Player Seeds (PS) produced in QUANTUM MOVES 2 and Random Seeds (RS). These seeds are subjected to an optimization algorithm running either on a computer cluster (GRAPE or SA) or in-game on individual player devices (Player GRAPE (PGRAPE)²). We term the combination of seed type and algorithm a *method*: most of the work concerns the absolute and relative performance of PGRAPE PS, GRAPE PS, GRAPE RS, and SA RS methods³. The joint set of player seeds and their own local optimizations are denoted $\text{PGRAPE PS} \cup \text{PS}$.

The crowdsourcing aspects in this work thus lie along two axes. The first lies in incorporating “human common sense” and decision making into the algorithmic loop, primarily in the seeding, that is otherwise elusive and/or difficult to implement programmatically. The second lies in the players donating their CPU clock cycles for optimizing seeds. This type of distributed computing is very valuable to the researcher in large scale studies⁴ since these iterations are “for free” and — if the solution is not yet at the local optimum it can simply be optimized further in the usual way.

Before giving a more quantitative presentation we briefly summarize some of the main results. Using different methods we map out the optimized $F(T)$ behavior near and below $T_{\min}^{F=0.99}$ and $T_{\min}^{F=0.999}$. The optimal control strategies for each problem are identified and their plurality and T -dependence are elucidated, cf. Sec. 3.2.1. Based on the aggregated results, a relative (numerical, landscape, and overall) complexity hierarchy among the problems are established. In terms of comparing the methods’ performances we consider different metrics and make the following observations. First, across all three problems, PGRAPE PS performed roughly on par with the best of the tested conventional methods (GRAPE RS or SA RS, depending). Second, GRAPE PS was the only method that yielded roughly optimal results in every problem (the rest failed separately in at least one). Third, results achieved by optimizing restricted sample sizes of PS were found to possess significant statistical advantages over doing the same with RS (we recover the “full scale” absolute results in the limit of sufficiently many samples). Further, picking only the most promising seeds (measured by fidelity) in this procedure significantly increases the performance for PS while the same is not found for RS. This is interpreted as PS having a higher probability of being in the green region(s) S_2 in Fig. 8.1. These findings support the hypothesis that player/human contributions may potentially be useful additions to quantum control at least for the studied problems.

² GRAPE and PGRAPE are identical algorithms, but the termination criteria of the latter are set by the player.

³ Note RS are only optimized on the cluster whereas PS are optimized on both cluster and player devices.

⁴ See also the Folding Home project at foldingathome.org.

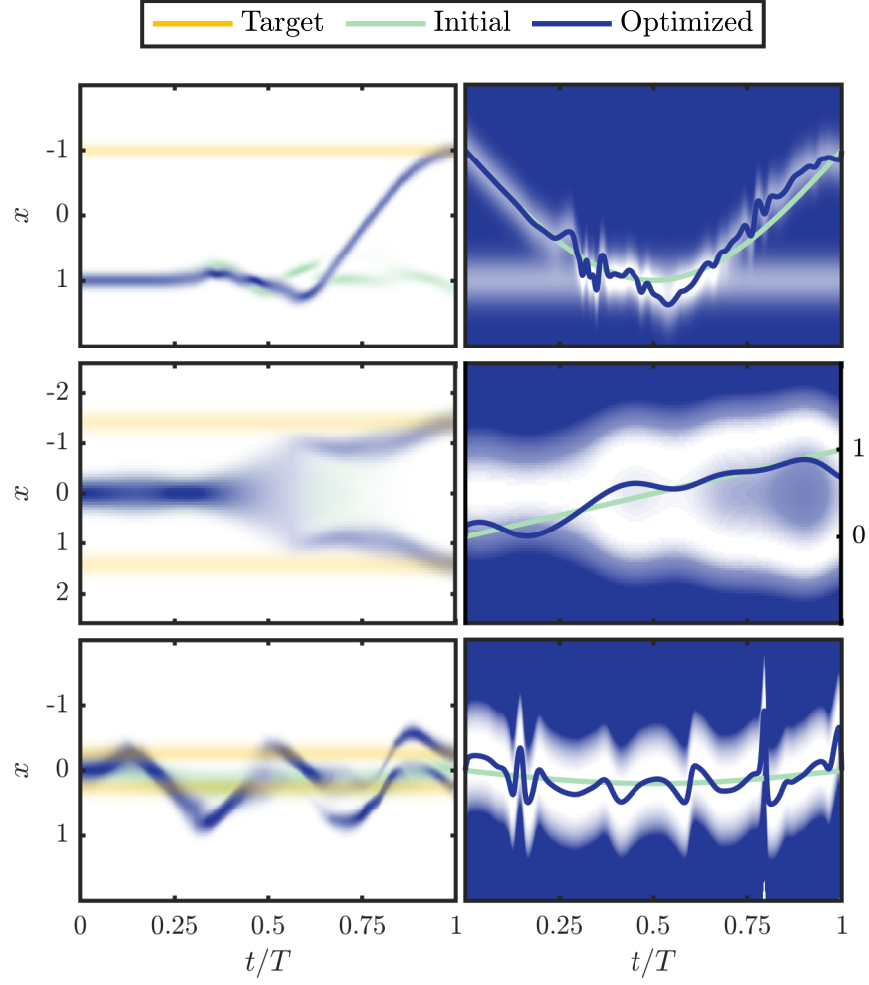


Figure 8.3: The three state transfer problems considered in the presented manuscript [5] and described in the text. Here, a few optimizations yielding $F \geq 0.99$ at large $T \gg T_{\min}^{F=0.99}$ are shown for illustrative purposes. The initial and optimized controls are shown with solid lines (right panels) and the target state and associated dynamics are visualized (left panels). The left panels show state densities and right panel colormap shows the combined potential (arbitrary units, blue represents large potential values) with same y -axis.

Top *Bring Home Water* with $V = V_{\text{tweezer}}(x_0(t), A(t)) + V_{\text{tweezer}}^{\text{static}}(1, -130)$ in Eq. (2.4) with $(x_0, A)(0) = (x_0, A)(T) = (-1, -130)$ and $w_0 = 0.5$.

Middle *Splitting* with $V = V_{\text{split}}(u(t))$ in Eq. (2.9) with $u(0) = 0$ and $u(T) = 1$. The control u does not have spatial units and the values instead are shown on the right axis.

Bottom *Shake Up* with $V = V_{\text{shake}}(u(t))$ in Eq. (2.10) with $u(0) = u(T) = 0$. One dimensional renditions of the potentials are also shown in Fig. 2.1. Units and more detailed descriptions are given in the manuscript [5].

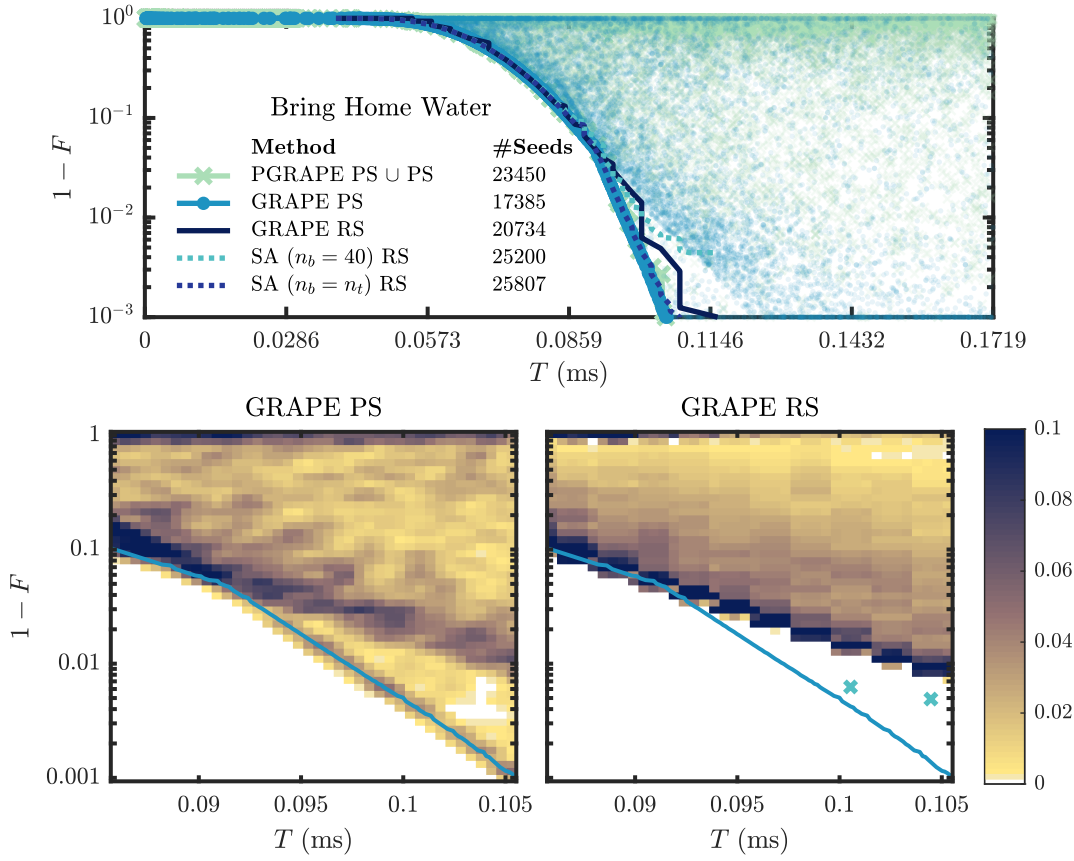


Figure 8.4: Adapted from Ref. [5]. Top: Optimizing using different methods (lower is better). Solid lines with (without) symbols show the best results obtained for individual methods with (without) player influence and the scattered translucent symbols show all individual solutions. A total of 536 players seeded on average ~ 32 solutions and optimized approximately 1/4 of these for ~ 131 iterations on average. Bottom: The colormaps show solution densities (each individual column is unit normalized) due to the GRAPE PS and GRAPE RS methods and the reference line shows the monotonically best GRAPE PS results from the top panel. The crosses mark single solutions at very low densities.

8.4 MANUSCRIPT RESULTS

The exposition given here is meant to convey the main results and is somewhat abridged — it does not contain all the granular details, technicalities, and observations contained in the rather extensive manuscript. Method descriptions are similarly more conveniently glossarized there for reference. For example, we shall focus primarily on RS and PS when applying the same gradient-based optimizer (GRAPE and PGRAPE, differing only in termination criteria). That is, we leave out detailed discussions on the derivative-free SA⁵ since it was only applied to RS — it is interesting in its own right but here we focus on an apples-to-apples comparison for the different seed types.

⁵ I give an in-depth explanation of the SA algorithm itself in the manuscript appendix.

8.4.1 Bring Home Water

In this single-particle task the goal is to pick up an atom, residing initially in the ground state of a static tweezer, and bring it back to ground state of a controlled tweezer at its origin point cf. Fig. 8.3. The control parameters are the position and amplitude, $V_{\text{tweezer}}(x_0(t), A(t))$.

The top of Fig. 8.4 shows the monotonically best optimization results for all methods. The landscape is seen to contain many local traps since many points lie above the empirically found best achievable $F(T)$. The best achievable $F(T)$ itself is piecewise exponential and evolves into two distinct and exponentially gapped branches. These are more clearly seen in bottom of Fig. 8.4 which shows the normalized solution densities using GRAPE with the two different seeding types PS and RS. It is observed that PS finds both branches while RS finds only the worse one.

Applying Density-based Spatial Clustering of Applications with Noise (DBSCAN) clustering [405] on GRAPE PS solutions near the two branches we find identify with each branch a distinct optimal control strategy and these are shown in Fig. 8.5. The very well-defined strategies are termed *front-swing* (cluster 0) and *back-swing* (cluster 1) because they initially place the tweezer in front of and behind the center of the atom, respectively. It is at first glance surprising that the back-swing strategy becomes dominant for all sufficiently large T since the atom travels an overall longer distance — it turns out that initially shifting the atom “backwards” and then appropriately overlapping the tweezers followingly provides a very large acceleration towards the target (the total potential gradient becomes very steep and positive) and this makes up for the longer distance.

A natural question remains: why does using RS with GRAPE prevent finding the back-swing, i.e. the exponentially best, optimal strategy whereas the same with PS does not? The answer is that the completely unstructured seeds (no correlation between subsequent control values) yields a vanishing probability that the atom will initially accelerate away from the target position. In other words, the seed is never placed close to the back-swing attractor in the control landscape and therefore cannot be found by local optimization. In Fig. 8.6 we introduce correlations into the seed⁶ by reducing the number of piecewise constant segments $M \leq N_t = T/\delta t$ cf. Sec. 3.4. The low values $M \sim \mathcal{O}(1)$ necessary to find the back-swing with RS is accompanied by a significant increase in poor solutions and, no matter the value of M , it is found that PS is always more efficient in this dataset.

⁶ The control itself is still parametrized as a piecewise constant with a value for each δt . The optimizer thus still explores the same overall parameter space but the seeds focus on narrower areas.

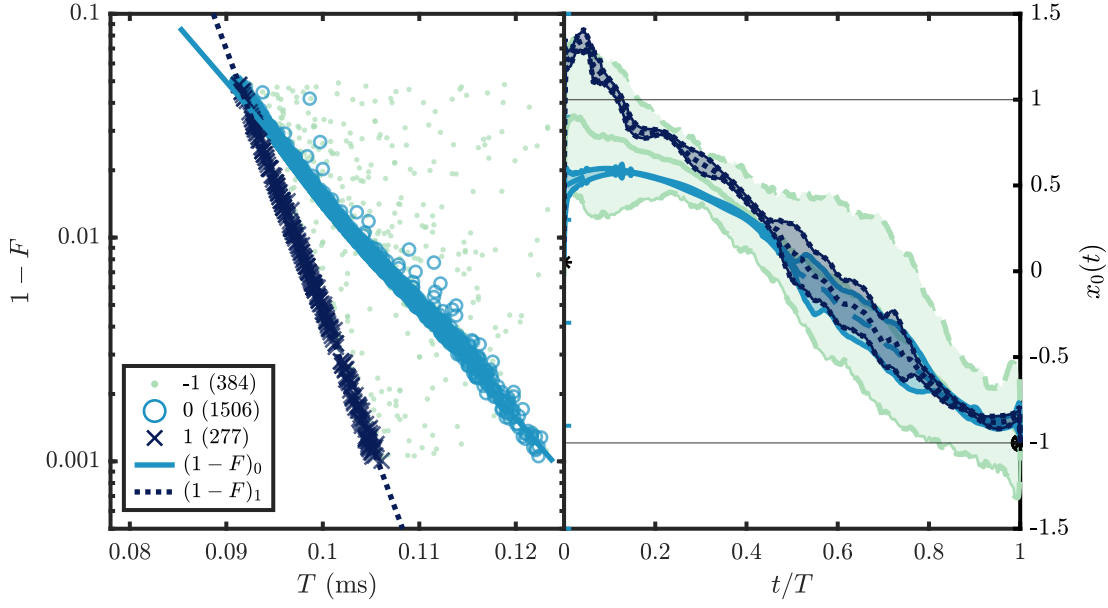


Figure 8.5: Adapted from Ref. [5]. Clustering results. Left: Optimal control strategy clusters (0,1) and their empirical multiplicities in $F(T)$ space. Each cluster member falls close to the respective exponential fit lines $(1-F)_{0(1)}$. Controls outside the clusters are denoted (-1). Right: The $x_0(t)$ control quantiles (25%, 50%, 75%) of the optimal strategies. The horizontal lines denote the atoms' initial position ($x_0 = 1$) and the target position ($x_0 = -1$).

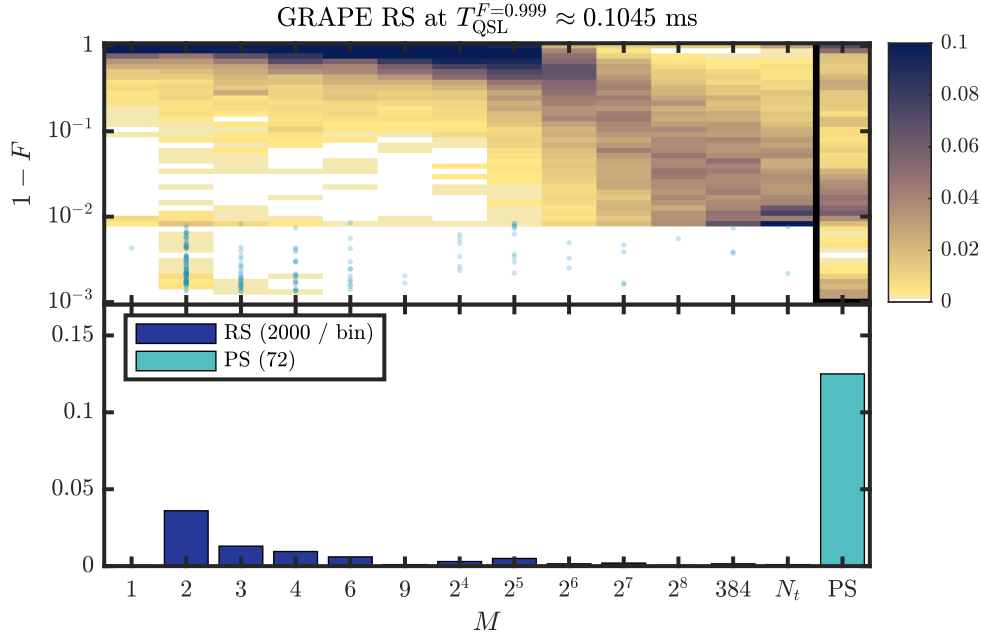


Figure 8.6: Adapted from Ref. [5]. GRAPE RS solution densities and histogram of solution fraction in the back-swing strategy when optimizing seeds with variable number of piecewise constant segments M . The dots denote single solutions in low density areas. The PS column shows corresponding results for GRAPE PS from Fig. 8.4.

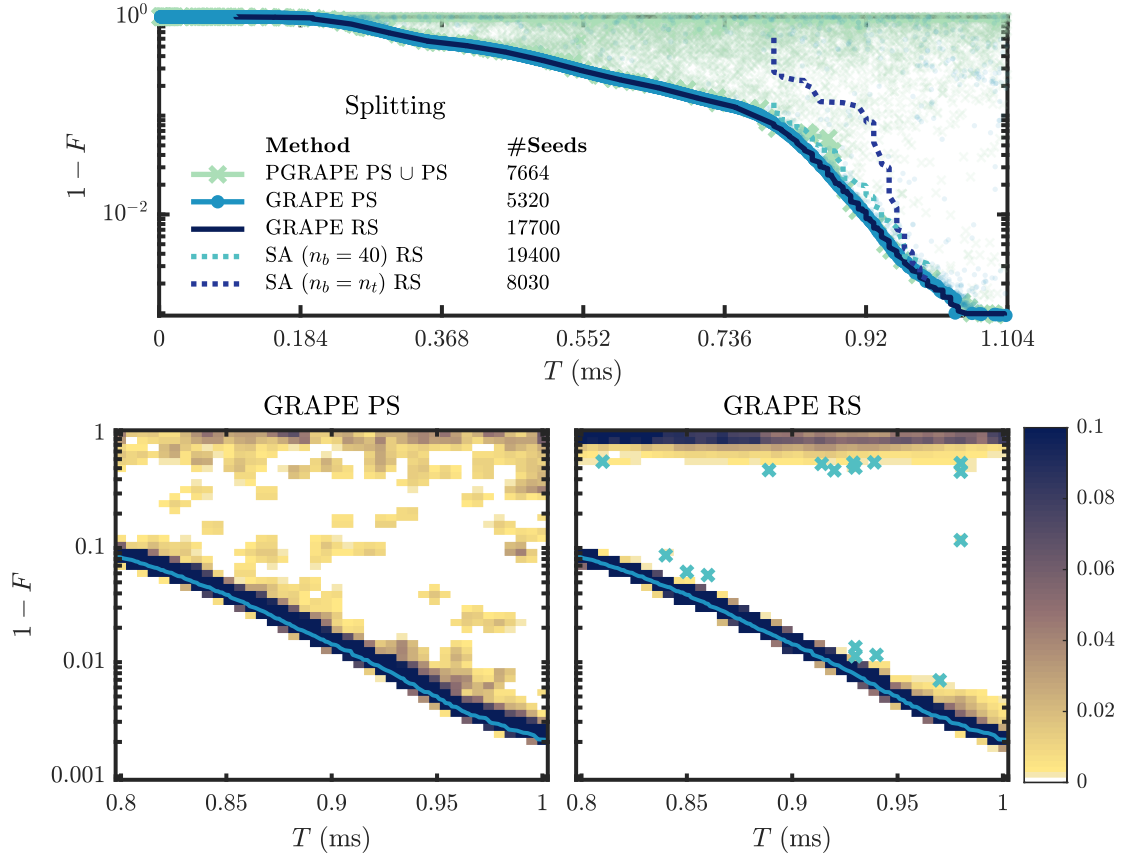


Figure 8.7: Same as Fig. 8.4, but for *Splitting*. A total of 193 players seeded on average ~ 28 solutions and optimized approximately $1/3$ of these for ~ 127 iterations on average.

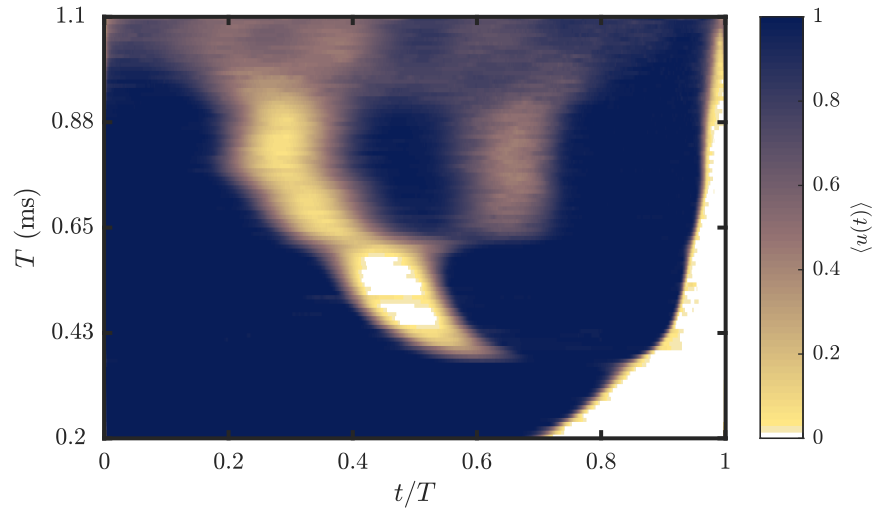


Figure 8.8: Adapted from Ref. [5]. Clustering results where a single optimal control strategy is found. The mean cluster control is a bang-bang structure that tapers off with T .

8.4.2 Splitting

The goal is to effect a meanfield BEC ground-to-ground state transfer, each state defined by a single-well and double-well configuration, respectively. That is, the BEC is split into two wave packets with the same phase by modulating the central barrier height ($\sim u(t)$) of the potential $V_{\text{split}}(u(t))$ cf. Fig. 8.3.

Figure 8.7 shows the optimization results and is seen to be simple and nearly trap-free⁷. Both seed types achieve the same empirically best $F(T)$ and the solution densities are roughly equally good with a slight advantage to PS.

Applying DBSCAN clustering on GRAPE PS solutions near the monotonically best ones we clearly⁸ identify a single optimal control strategy. Figure 8.8 shows the mean control over all the cluster members. At low T it is characterized by a bang-bang structure corresponding to switching between maximal and minimal central barrier heights: the initial maximal value provides the steepest acceleration of the condensate into the separate wells. With increasing T the bangs become increasingly blurry.

8.4.3 Shake Up

The task is to drive a meanfield BEC from the ground state into the first excited state. This is achieved by shaking the anharmonic trap center given by the control u cf. Fig. 8.3.

The optimization results in Fig. 8.9 exhibit staircase-like plateaus that partially obscure the empirically best achieved $F(T)$. This is also seen in the solution density which is of high variance and the high-density concentration is shifted mostly away from the best achieved results. The seed types yield somewhat similar results when applying GRAPE. However, for this dataset we do observe that PS outperforms in regions $T \leq 0.85$ ms and $T \geq 0.98$ ms and RS outperforms in between. This is different from the two previous problems where GRAPE PS was always better or as good as GRAPE RS for all T . We therefore join the two sets in the clustering procedure for identifying the optimal control strategies.

The clustering was unsuccessful when applied to the controls directly in the time domain, but transforming into the frame following their individually induced center-of-mass motion, $\langle \hat{x}(t) \rangle$, we identify

⁷ The green symbols lying above the lines are due to the players terminating the optimization prematurely.

⁸ Roughly 99% of these 3625 solutions belonged to the identified cluster.

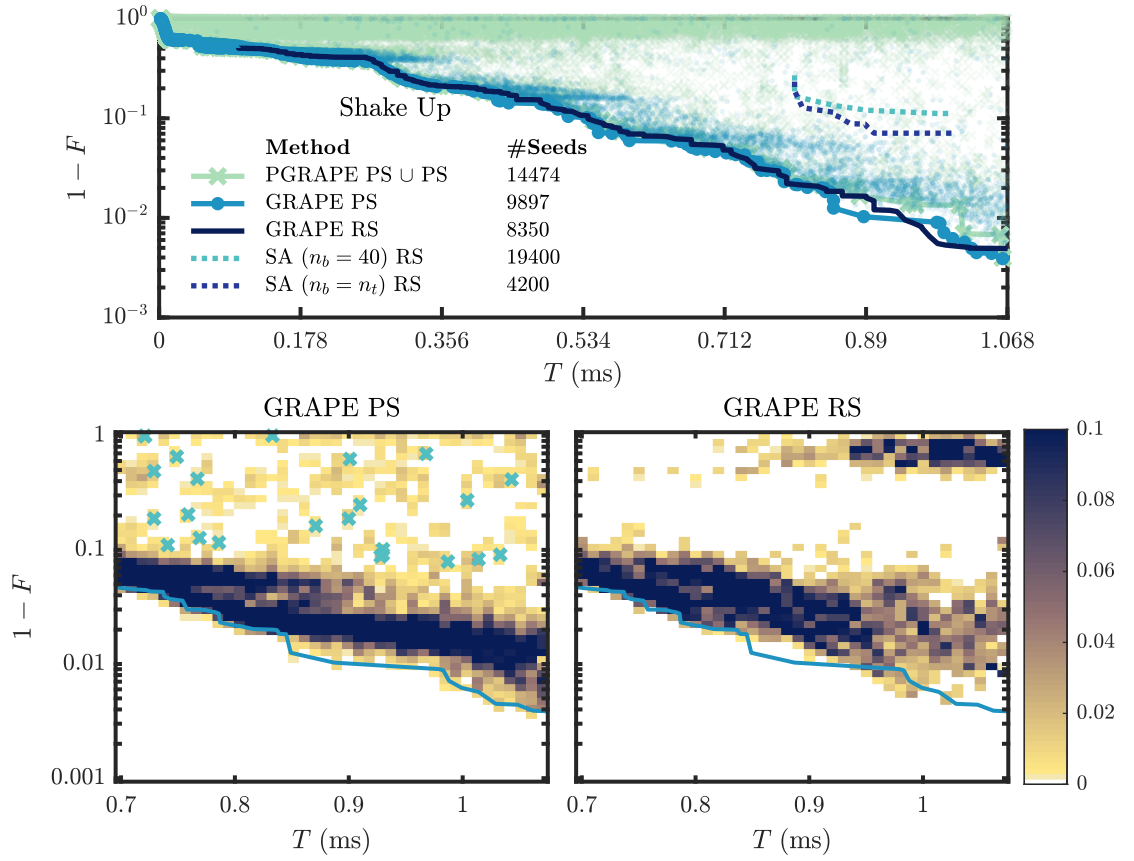


Figure 8.9: Same as Fig. 8.4, but for *Shake Up*. A total of 266 players seeded on average ~ 37 solutions and optimized approximately $1/3$ of these for ~ 116 iterations on average.

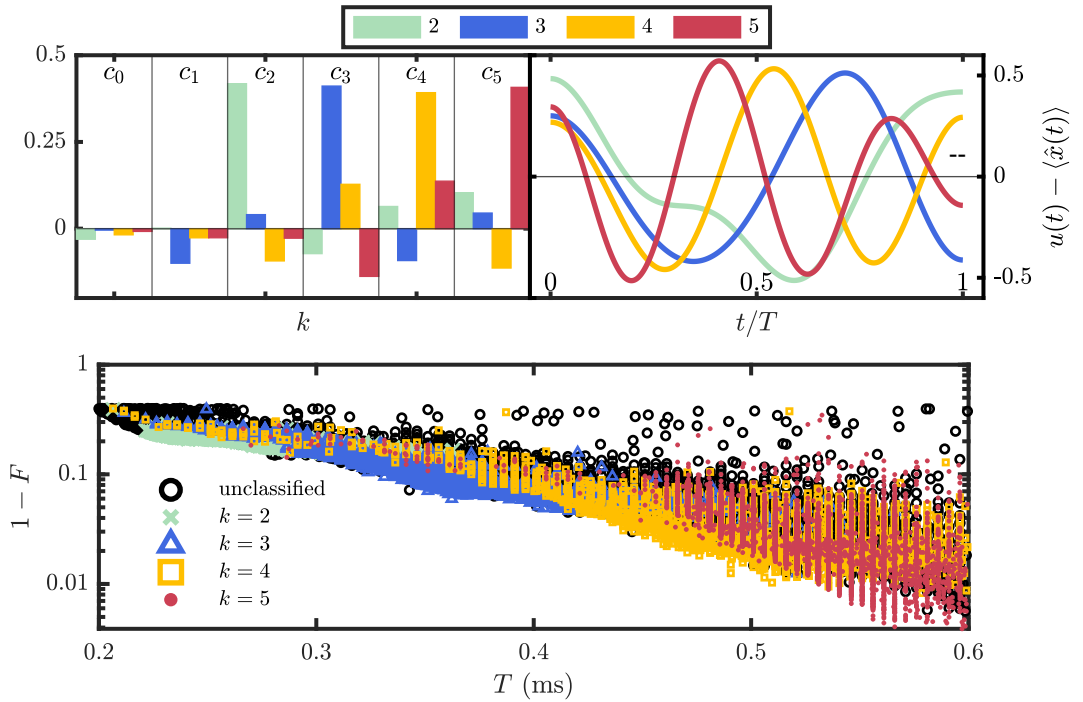


Figure 8.10: Adapted from Ref. [5]. Clustering results according to Eq. (8.1) with four identified optimal control strategies labeled corresponding to the dominant frequency component $k = 2, 3, 4, 5$, correspondingly $N_k = 1, \frac{3}{2}, 2, \frac{5}{2}$ oscillations. The cluster means are shown in frequency space (top left) and in real space (top right) both in relation to the center-of-mass frame. The distribution in $1 - F$ space (bottom) shows a hierarchy of interleaved optimal strategies (see Ref. [5] for translucent version).

multiple strategies characterized by low-frequency cosine components. Specifically, all the relative/transformed controls are decomposed as

$$c_k[u(t)] = \frac{1}{T} \int_0^T (u(t) - \langle \hat{x}(t) \rangle) \cos(\pi k t / T) dt, \quad (8.1)$$

where $k = 0, 1, 2, \dots$ correspond to half-integer oscillation numbers $N_k = k/2$ and we perform the clustering on all resulting $c = (c_0, c_1, \dots, c_5)$ vectors. The results⁹ are shown in Fig. 8.10. A total of 4 strategies are found and each is characterized by a dominant component k (number of oscillations N_k correspondingly) and this is clearly reflected in the time domain. The strategies take turns being the best one, covering roughly equally wide duration intervals, with increasing T leading to more oscillations being desirable. However, even outside the region where they are the best choice, each strategy remains a sizeable attractor in the landscape and thus obscures the current globally best strategy. This leads to the previously observed plateaus.

8.4.4 Summary and Comparison of the Problems

The aggregate results presented above each contains on the order of thousands to tens of thousands solutions. In this highly dense sampling limit, it was found in these datasets that **GRAPE PS** is always among the best methods with a short exceptional window in *Shake Up* where **GRAPE RS** is better. Additionally it is observed that the best results of **PGRAPE PS** $\cup$ **PS**, which is entirely due to the players and their own device-local optimizations, is roughly on par with the overall best results.

In analyzing the results we found the following characteristic numerical¹⁰, landscape, and overall complexity hierarchies.

Splitting is a **BEC** problem and the numerical complexity is thus high.

It has a very simple, nearly trap-free control landscape with a single optimal strategy that was found by all algorithms — so simple in fact that even the undiscussed **SA** did so despite being excessively numerically ineffective for **BEC** problems. Thus, it has the lowest overall complexity.

Bring Home Water is a single-particle problem and the numerical complexity is thus low. It has a relatively high landscape complexity with many scattered local traps and two exponentially gapped optimal control strategies. The best one was much harder to find and not all methods were able to do so. Thus, it has the middlemost overall complexity.

Shake Up is a **BEC** problem and the numerical complexity is thus high.

It has a very high landscape complexity, also with many local traps,

⁹ Note that $T^{-1} \int_0^T \cos(\pi k' t / T) \cos(\pi k t / T) dt = \delta_{k,k'} / 2$ for integers k and k' .

¹⁰ "High" and "low" numerical complexity is here understood only in relative terms.

and exhibited four much more intricate and interleaved optimal strategies related only in a nontrivial parametric frequency space. The best strategy at T was partially obscured by the others and this lead to all methods struggling to map out the best achievable $F(T)$. Thus, it is the highest overall complexity.

All problems are characterized by (piecewise) exponential behavior in the best empirically achieved $F(T)$.

8.4.5 Statistical Performance

The aggregate results presented in Secs. 8.4.1–8.4.3 contained on the order of $\mathcal{O}(10^3) - \mathcal{O}(10^4)$ individual solutions for each method. We now examine the statistical performance of all methods¹¹ as the number of solutions are reduced to increasingly sparse limits. This could equivalently represent a situation with higher numerical difficulty or limited optimization resources — both interpretations lead to fewer seeds optimized within a fixed time frame.

To this end we follow the procedure¹² detailed in the manuscript where a fitted mean $\langle T_{\min}^{\text{fit}} \rangle$ and a success probability $0 \leq \mathcal{P} \leq 1$ are calculated for different sample sizes N_{samples} . The results of the sampling procedure is shown in Fig. 8.11 as a function of N_{samples} . In the dense limit $N_{\text{samples}} \sim \mathcal{O}(10^4)$ all gradient-based methods are roughly equally performant and coincides with the aggregate results in the above sections. When N_{samples} is decreased, however, the methods' performances drop off one by one. In this way it is seen that **PGRAPE PS** is mostly inferior to **GRAPE RS**. Players evidently terminate the optimization prematurely on average. Yet, **GRAPE PS** (the fully optimized/realized player seeds) display notable statistical superiority to **GRAPE RS** in the increasingly sparse sampling limit.

The dataset also includes results when sampling only from the best (measured by fidelity) aggregate results for each seed type and this is denoted by the prefix **pr** (“preselection”). We also define **po** $\equiv$ **PGRAPE PS** as solutions partially (or even fully) optimized by players that can then conditionally be optimized further as “seeds” on the cluster. We observe that the **pr** heuristic greatly improves the results for **po** and **PS** while the same is not observed for **RS**. This suggests that the structure of **PS** is preferable to the lack of structure in **RS**. That is, they are much more likely to be in the green region S_2 of Fig. 8.1 and this supports the baseline hypothesis that player contributions may be useful in uncovering the optimal control strategies. Additionally, the respective

¹¹ The manuscript also shows statistical run time comparisons for the methods.

¹² Briefly, we draw a number of samples, N_{samples} , fit an exponential to the monotonically best $\log_{10} 1 - F$ values, and locate the $T_{\min}^{\text{fit}}$ such that the fit yields $F = 0.99$. If the T is within an acceptable region (interpolation) it is a success and otherwise a failure (extrapolation). The procedure is repeated 1000 times and we record the mean $\langle T_{\min}^{\text{fit}} \rangle$ and the success fraction $\mathcal{P}$.

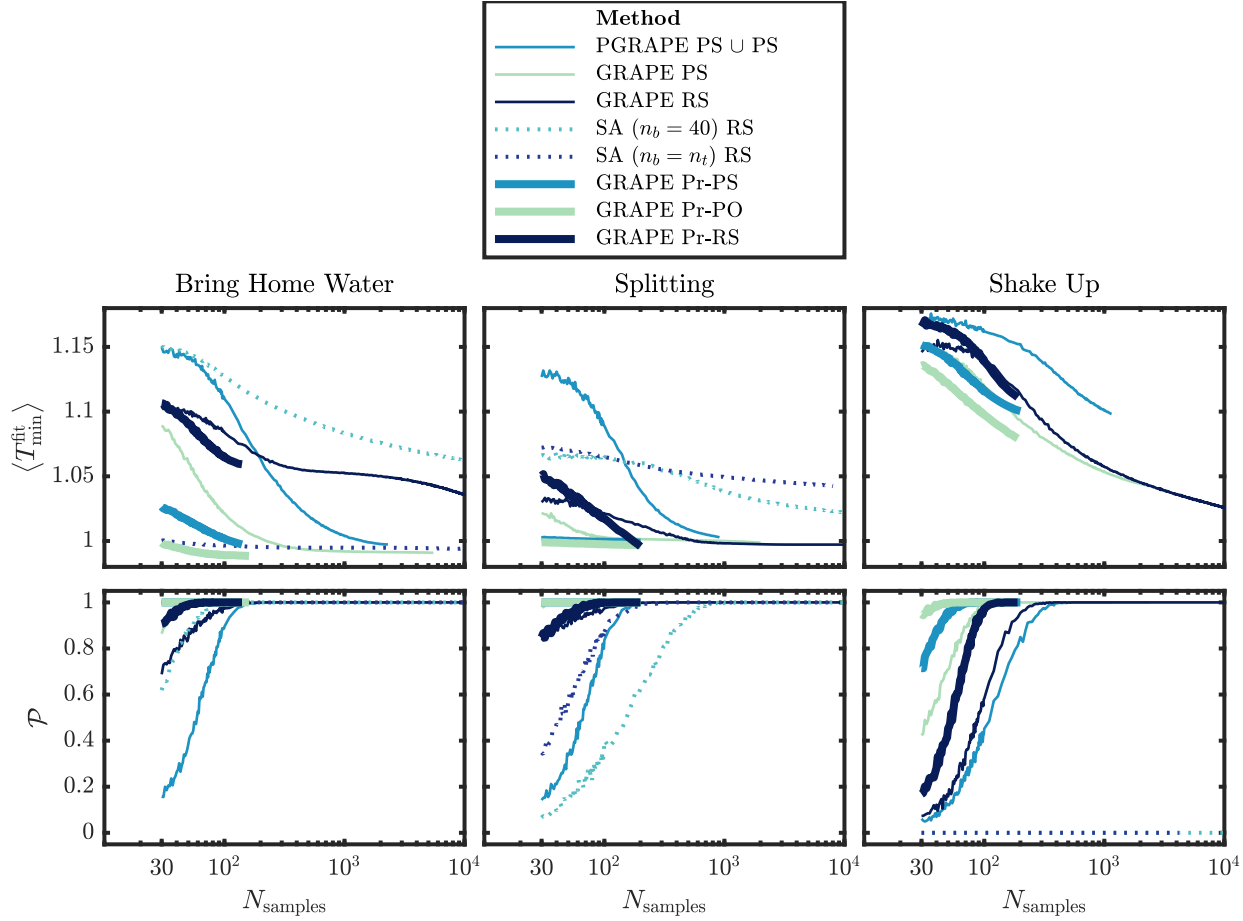


Figure 8.11: Adapted from Ref. [5]. Statistical performance as a function of sample size in each problem (column-wise) in terms of $\langle T_{\min}^{\text{fit}} \rangle$ (top row) and $\mathcal{P}$ (bottom row). Good statistical performance requires simultaneously low $\langle T_{\min}^{\text{fit}} \rangle$ and high $\mathcal{P}$.

performance difference between GRAPE pr-PO and GRAPE pr-PS roughly signify the (evidently nonzero) gain in first letting players partially optimize the seeds.

The observations made here and in previous sections suggest that crowdsourcing the solution of the studied problems was successful.

8.5 PERSPECTIVES

We now turn to some of the broader perspectives and future possible directions for similar work.

8.5.1 Disclaimers

First it is pointed out that the results of the manuscript do not claim to guarantee that players or citizen scientists can or will outperform algorithms on their own. Players simply cannot be feasibly expected

to possess neither the manual dexterity nor inherent skill to “magically know” how to produce *exactly* correct $F \geq 0.99$ dynamics. Such expectations are also not part of the hypothesis where the fine-grained level of control is achieved by the optimization algorithms.

The results also do not claim to guarantee that players *together* with the concrete algorithms will always outperform results achieved by fully expert-driven optimizations. This is regardless of (1) whether the latter uses the same or other past, present, or future algorithms, and (2) the choice of problem, including the ones considered here. Especially, domain specific knowledge is, of course, used by experts all the time in devising increasingly sophisticated seeding strategies and methodological developments to target promising regions of control space. This sentiment is hopefully evident from the discussions extensively scattered throughout this thesis. Any control problem is simple when viewed in an appropriate parametrization.

8.5.2 Perceived Significance of the Results

What the results *do* suggest is that human input can be a useful addition to the established conventional work flows. For example, the difficulty in choosing a good seeding strategy can nearly always be reduced to a nontrivial exploration vs. exploitation trade-off. If the expert’s computational resources are restricted and no “intuitive” solution to a given problem is known, one could imagine crowdsourcing the initial exploration stages [81] (seeding, optimization, or both) which could then guide the expert in refined followup exploitation stages.

As a concrete example, the identified optimal control strategies in the *Shake Up* problem were quite involved and it is virtually certain that we did not yet find the ultimately best achievable $F(T)$ — future optimizations could now search in and around the much narrower and relevant parts control space. This idea is in itself not exclusive to player seeds and could also be envisioned for any other type of seed as was also suggested in discussing the results of Ref. [4]. However, in the problems studied here we observed indications that player seeds could be distinctly useful because (1) they displayed several statistical advantages over fully random seeds, and (2) they provided separate datasets — free of potential expert biases — and in *Shake Up* this complementarity yielded enhanced collective results. Additionally, similar strategy discovery by players has also been reported in the Foldit [406] and even algorithmitized in the EteRNA [374] citizen science projects.

It is further remarked that, although a complexity hierarchy among the problems was established, they are all sufficiently simple enough

that large scale studies¹³ were feasible. Approximately optimal solutions in terms of absolute results were found by just casting a wide enough net, aggregately on the order of some/ten/hundred thousands of individually optimized solutions in each problem. Clearly, the potential relative gain in absolute results from these methodologies is much more bountiful if they can be successfully extrapolated to problems of, e.g., significant numerical complexity where such an ultra-dense sampling is not feasible (i.e. only sparse sampling in the sense of Fig. 8.11 is possible). This scenario could, for example, be realized in the many-body regime of Ch. 7 and similar. Such extrapolations remain speculative until comprehensively studied.

8.5.3 Generalizability — Crowdsourcing

Although the explicit nature of the “common sense” aspect was not analyzed, one reasonable conjecture is that the QUANTUM MOVES 2 interface, which graphically displays the evolution of $|\psi|^2 + V$, facilitates an analogy to a classical sloshing liquid in an open container. One possible critique about methodological generalizability is then that the players would perhaps fail to some extent, on the same problems, if the interface instead displayed the momentum space (Fourier) representation $\mathcal{F}[|\psi|^2 + V]$ or similar as a “common sense” analogy may be less obvious to the players. However, it should be recalled that the first order of business, in any given problem, for a domain expert lies exactly in finding an appropriate *physical* representation (e.g. basis states, transforming in to rotating frames, etc.) through the lens of which the problem becomes as simple as possible. Thus, it is unfair to discount the methodology solely on the basis of a clearly poor interface representation because the same could be argued for the physical representation (choosing an “equivalently” poor representation) — it is ultimately the domain expert who is in charge of designing *both* and there seems to be no reason why they would intentionally hamper one and not the other¹⁴.

There are, however, other quite valid and nonpathological related questions relevant for surmising if any indicated advantages from human input can be maintained in general problems.

First, it may not be clear *what* constitutes an appropriate interface — this includes the set of features, UX and UI, etc., and the general design difficulty hereof was already mentioned in Sec. 8.2. For example, as far as quantum optimal control is concerned, the graphical representation of a given task must be adapted when considering, e.g., more than one

¹³ Incidentally, one possibility afforded by this type of research-relevant setting is to try and systematically map out human qualities in problem solving. Capturing these in developing cognitive models may be useful for advancing AI [389]. Other recent work in our group [407] on hybrid intelligence co-creativity is remarked in passing.

¹⁴ Outside of explicitly studying the effect of doing so.

spatial dimension, entangled N -body systems, decoherence/mixed-states, and spin/generalized degrees of freedom. A subtle difficulty lies in that a sensible representation for the domain expert, e.g. in terms of certain expectation values, may or may not be sensible for the nonexpert. Nevertheless, it was observed in other work from our group [81] that human input achieved quantitatively comparable, but qualitatively different, results to conventional optimization even in a black-box-like scenario, i.e. one where the problem visualization is abstract and dynamical analogies are absent.

Second, there is an obvious buy-in cost of designing, building, deploying, and maintaining these nontrivial interfaces. That is, the gain in crowdsourcing is in reality offset by opportunity losses. For example, with the same monetary and time budget, one could instead have invested in more compute and optimized orders of magnitude more random seeds or similar. When the problems are sufficiently simple, as those studied here, this would very likely have yielded as good or better absolute results compared to the players', offsetting any potential per-seed relative performance differences. There are, however, problems that are so complex that simply adding more brute force compute yields diminishing returns, e.g. in protein folding, and it seems reasonable to suspect that these may also exist within the quantum control arena. Instead of tackling these highly complex problems directly, the work here establishes that a crowdsourcing approach can be effective in the opposite limit — which is much more amenable to methodological comparisons — to begin with. In broader terms, the steep buy-in cost of establishing an interface is significantly more palatable if one envisions it as an evolving platform on which researchers can continually feed in new problems to be solved fully or partially by the players.

In conclusion, the generalization of this and similar crowdsourcing methodologies may be enticing and indeed possible, but it is by no means an automatic guarantee of equal or better results than usual. Each instance must be evaluated empirically on a case-by-base basis and comprehensively mapping this out remains for future work.

8.5.4 Generalizability — Interfaces

Human-in-the-loop ideas do not discriminate specifically against domain experts — they are also humans, after all. Thus, novel interfaces can also be envisioned for use by experts or semi-experts and these can be built from the ground or repurposed from existing ones such as those stemming from citizen science projects [408]. The design goals of such interfaces will obviously be quite different than those intended for nonexperts. These could, e.g., aid in solving the same problems but with an expanded set of features. They could also be used in entirely separate settings, e.g. exploring problems from a different

angle, streamlining work flow, and generally be a useful addition to the toolbox of professional scientists.

As an example of this, the following Ch. 9 describes one such interface, QUANTUM COMPOSER, that was initially inspired from our citizen science games.

MANUSCRIPT [5] – CROWDSOURCING HUMAN COMMON SENSE FOR QUANTUM CONTROL

Published in Physical Review Research.

22 pages.

AUTHOR CONTRIBUTIONS

This project was a collaborative effort between theory and the ScienceAtHome developmental unit and had many contributors at many different stages.

In the development phase, I had the overview of decisions, processes, time lines, and alignment between the scientific and developmental units, both on a higher level conceptual scale and at a lower level technical scale. I was responsible, along my fellow PhD student Shaeema Zaman, for administrative tasks such as planning and running meetings of scientific, and occasionally developmental, nature. Based on these meetings I continually planned, developed, and implemented the so-called *Bindings* layer that both “binds” the QEngine (C++) to the game development platform (C#) and exposes or extends functionality to realize our goal with the game, while making it effortless for the game developers to use — with whom I routinely had day-to-day meetings — by hiding the complexity behind simple function calls. For this part of the work, I had significant help from Jakub Czarkowski who was a masters’ student at the time. I also ended up being fully involved in an accompanying learning track (not associated with this publication) which I both reviewed several times and contributed to with a significant amount of work.

In the analysis phase, Jakub Czarkowski performed most of the cluster optimizations first under my supervision and later also Miroslav Gajdacz. Miroslav Gajdacz performed the clustering and initial solution density analyses. Many authors contributed to the analyses of the majority of the results, and all authors participated in their interpretations through scientific discussions.

In the writing phase, I authored the manuscript and figures with scientific feedback and corrections primarily from Miroslav Gajdacz, Shaeema Zaman, Carrie Weidner, Klaus Mølmer, and Jacob Sherson. There are a few exception (to which I provided scientific feedback and corrections) as follows: 1) the first paragraph in the introduction was authored by Jacob Sherson and Janet Rafner, 2) the five posed citizen science questions (Q1-Q5) in the introduction were authored by Jacob Sherson, 3) the first paragraph beginning on p. 17, column 2, regarding other contexts than pure state transfer was authored by Miroslav Gajdacz, and 4) the paragraph on *Quantum Moves 1* in the introduction was authored by Jacob Sherson and Klaus Mølmer (that work was before my time).

Note: The manuscript [6] reported on in this chapter is topically unique from the others. In particular, the content therein is not suitable for same the presentational style given elsewhere in this thesis. On the basis of tonal cohesiveness, this chapter is therefore relatively short and provides instead primarily a higher level presentation.

QUANTUM COMPOSER (or just **COMPOSER** for short) is a visually programmable and interactive node-based tool for quantum mechanical numerical simulation and optimal control. It is a non-commercial distributable runtime application written in native C++ built on top and in extension of the QENGINE. In general, it features a lowered barrier to entry compared with conventional, pure text-based programming. Text-based programming can be both time consuming or cumbersome to write and establish build pipelines for — even in small projects — and may have to be recompiled and linked (depending on language) after making modifications to the source code. Additionally, results from text-based programs typically need to be visualized or analyzed for interpretation, possibly in a separate language. QUANTUM COMPOSER incorporates all these steps without the need for recompilation/linking and has certain similarities to the LABVIEW programming language for those familiar.

As of writing, COMPOSER supports a subset of the physical models implemented in the QENGINE, namely the one-dimensional Schrödinger and GP (meanfield BEC) systems. The application allows studying these types of first quantized systems where fundamental quantities of interest such as wave functions ψ , densities $|\psi|^2$, arbitrary linear combinations of states $\psi = \sum c_i \cdot \chi_i$, state comparisons $\langle \phi | \psi \rangle$, state operations $\hat{O} \cdot \psi$, density integrals $\int_a^b |\psi|^2 dx$, observable expectation values $\langle \hat{O} \rangle_\psi$ and variances $\text{var}_\psi \hat{O}$, arbitrary trapping potential geometries $V = \sum k_i \cdot V_i$, spectral information $\{E_i, \phi_i\}$, etc., can be calculated and visualized in real-time for both static and dynamic (i.e. time-independent and time-dependent) scenarios. A relatively expansive documentation can be found [here](#).

The strength of COMPOSER lies in the abstraction of all these concepts, and more, into predefined “atomic” node units. A simple example is shown in Fig. 9.1. Each individual node is defined by a certain set of inputs, interactable content, internal computations, and a set of outputs. Nodes are in this sense similar to functions in a functional

programming paradigm and each correspond to a number of actually compiled lines of code. Nodes with matching input and output can be connected to form a *flow-based* simulation — data “flows” from the sender’s outputs to the (potentially multiple) receiver’s inputs where it is used for the internal computations and producing output and the process continues: any change in the connection structure or data triggers a cascading sequential recomputation of all downstream nodes.

A particular set nodes, their content, and connections between them is called a *flowscene* (these can be saved and loaded to small .json files) and defines a numerical quantum mechanical simulation of almost arbitrary complexity courtesy of the emergent combinatorial possibilities stemming from the extensive list of different nodes. This (potential) complexity, however, can be further abstracted away by the user assembling one or more *Scopes* which are customizable groupings of nodes with collective input/outputs (i.e. a “superfunction” composed of sequential atomic functions). A *Scope* can be “collapsed” (or “folded”) to hide the details and can be used equivalently to the atomic node units. An advanced example where many of the details have been abstracted away is shown Fig. 9.2.

9.1 CURRENT AREAS OF APPLICATION

With these core features, QUANTUM COMPOSER can be used in a variety of different contexts ranging from educational to research settings.

9.1.1 Education

The roles of visualization and simulation tools are recognized [409–418] as productive supplements to conventional textbook-based classroom activities. Several applications have been developed and successfully applied in this regard [414–416] and these typically entail focused scenarios where the user can interact with a fixed collection of visual components. The scenarios and interactions are preprogrammed in text-based source code and compiled into small standalone applications by the respective service providers. What distinguishes QUANTUM COMPOSER compared to these is the flexibility of having a single unified framework that allows broad possibilities for intuitive on-the-fly adaptations suited to the particular needs of the end user which could be e.g. a lecturer, instructor, or student. At the same time, the ability to abstract functionality away into *Scopes* enables the same type of focused scenario as offered by the alternatives. It provides immediate and visual feedback mechanisms that support, among others, iterative inquiry-based learning.

COMPOSER has been used (to varying degrees) throughout multiple university courses in Aarhus and Europe on both the undergradu-

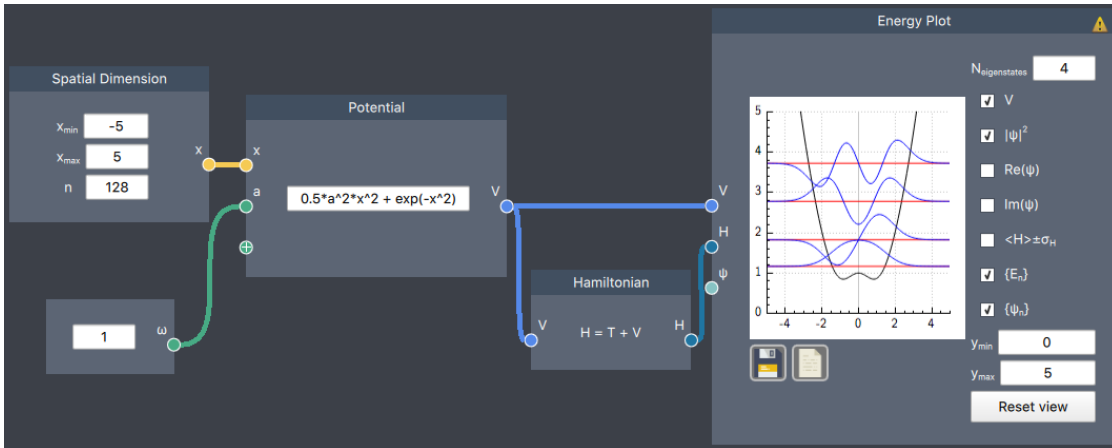


Figure 9.1: Simple flowscene that calculates and visualizes the spectrum of a given potential. Nodes have inputs (outputs) on the left (right) side shown by the round anchors. An input and output can be connected if they are of the same type (represented by color). The Potential node contains an input field for parsing potential expressions, here chosen to be $V(x) = \frac{1}{2}a^2x^2 + \exp(-x^2)$, where the spatial coordinate x and scalar a values are provided by other nodes. The output from the Scalar node has been re-labeled by the user to ω since it represents a frequency. The output from the Potential is connected to the Hamiltonian and the Energy Plot nodes where the potential and corresponding four lowest lying eigenstates and eigenenergies are plotted.

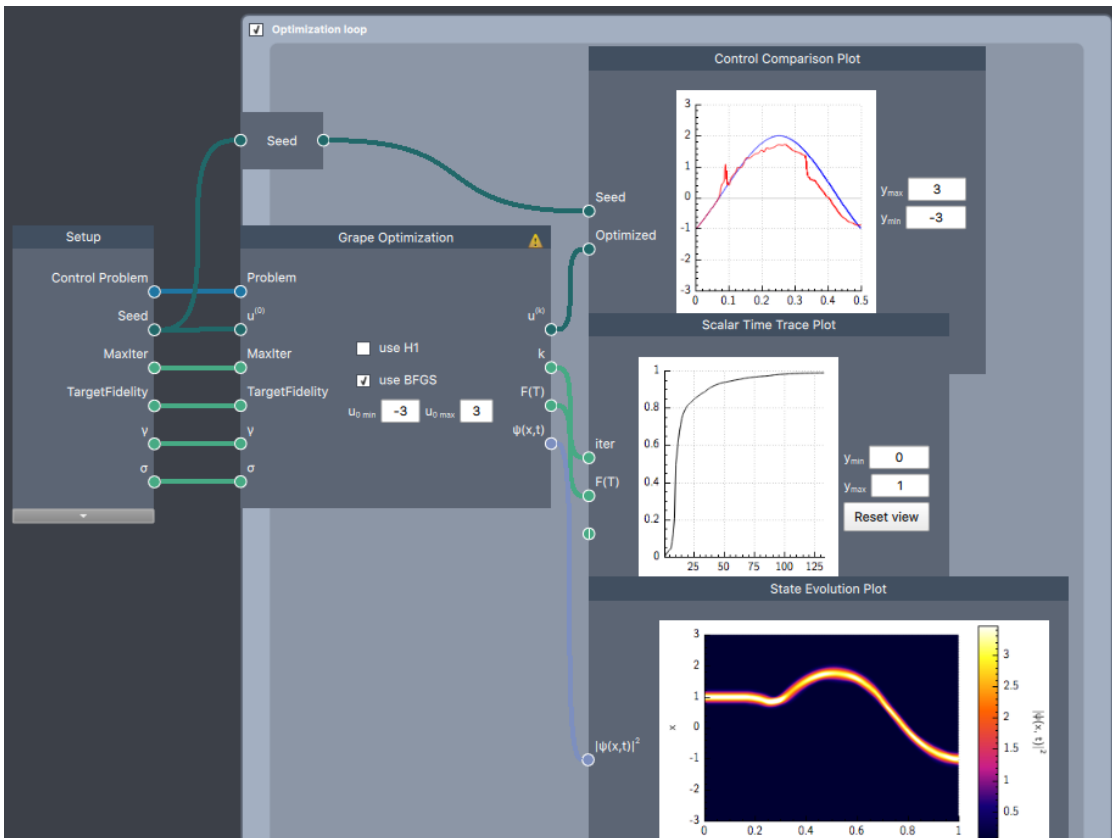


Figure 9.2: Advanced flowscene solving the *Bring Home Water* state transfer problem (compare Fig. 8.3). The (collapsed) setup Scope outputs relevant variables to the optimization loop Scope which then runs and visualizes data from the optimization in real-time. Both Scope contents have been programmed by the user.

ate and graduate level. It has been used for showcasing both introductory quantum mechanical phenomena (superposition, dynamical time-evolution, expectation values, ...) and advanced topics such as quantum control. It has also been used in internal workshops for new students in our group [6]. We have developed multiple modularized, distributable, and modifiable *exercises* on several topics and phenomena with accompanying descriptions, questions, and flowscenes targeted at different audiences. The exercises vary in nature and types of user engagement, including tightly focused cause-effect scenarios (e.g. “what happens to the spectrum as frequency ω is increased in Fig. 9.1?”) to open-ended questions (e.g. “estimate the lowest possible $T_{\min}^{F=0.99}$ for the state transfer problem in Fig. 9.2 — what characterizes the solutions and the resulting dynamics?”) and may come as fully preprogrammed flowscenes or ask the student to program it in various degrees of step-by-step [6]. The balance of these depend entirely on the intended audience and learning goals.

Outside of Ref. [6] we have, as of writing, published a smaller QUANTUM COMPOSER work related to student learning [8] and a similar much more expansive and quantitative manuscript is in preparation.

9.1.2 *Research*

We have internally found QUANTUM COMPOSER a welcome addition to the toolbox throughout various stages of our scientific research projects, for instance in precursor studies and solution analysis. While it is not, currently, designed to be the primary or sole means for producing data in a typical publication scenario it has been applied as a *research-assisting* instrument that draws on its strengths. For example, I used it in Ref. [2] (reported on in Ch. 6) to study and verify certain aspects of the control problem as follows.

1. Some of the controls had a non-trivial effect on the potential shape. COMPOSER was used to better understand their relationship and this was useful in designing appropriate seeding strategies.
2. The results of the work were benchmarked on an earlier paper and COMPOSER was used to verify that their spectra agreed — and thus that the control problem was correctly implemented.
3. The problem consists in merging a pair of unit cells and it is a central assumption that these pairs remain “unmixed”. COMPOSER was used to find good control boundaries such that this would always be satisfied.
4. COMPOSER was used to verify that the conversion between SI and non-dimensionalized numerical units was correct.

I also used COMPOSER for similar types of tasks in Ref. [5] (reported on in Ch. 8). In particular, the reported estimated timescales for dynamics (Sec.V.B in the manuscript) were obtain in this way.

Outside of use as an aid in concrete publications we have used QUANTUM COMPOSER on innumerable occasions for “quick and dirty” small scale studies or projects that would otherwise be much less feasible time investments. The experimentalist team has, for example, shown in *rough numerical estimations* that it is possible for to create a “plug” in our optical lattices using a classically *attractive* Gaussian tweezer profile (i.e. by shifting transitions out of resonance). In their words [6],

“... it allows us as experimental researchers to quickly simulate complex scenarios that move beyond simple back-of-the-envelope calculations without the [direct] use of complex numerical software [the underlying QENGINE]. Indeed, we have experienced that what is shown here could feasibly be implemented and understood by an advanced Bachelor’s student doing a project in our lab.”

and similar examples can be found in our theoretical branch.

It is remarked in passing that, as also mentioned in Ch. 8, COMPOSER is a direct descendant from applications that repurposed and iterated on our research-oriented citizen science game interfaces.

9.2 FUTURE CONSIDERATIONS

The purposes and applications of the tool are therefore many-fold. It is also, however, somewhat narrow in its current scope as it was initially conceived to cover the needs and research directions of our own group. In gaining a broader appeal it would be tremendously useful to expand the physical models supported by the program. Most prominently, following near term releases will hopefully include the possibility of treating spin-half(s) and general low-dimensional finite systems. In addition to new computational nodes for associated operators etc. this would also entail new nodes for visualization e.g. Bloch spheres, spin correlation functions, and so on. This would open up significantly more didactic possibilities as well as research-relevant studies since an innumerable amount of, e.g., quantum control deals with these types of systems. The latter includes for example the study of spin-chain transport (spintronics), NMR systems, NV-centers, superconducting qubits, and so on.

An ambition in this regard would be to transform QUANTUM COMPOSER into a research-enabling standalone application — rather than

just a research-*assisting* one — where full, high-quality, publication-ready results can feasibly be produced over multiple research areas and types of research questions. This medium-to-long term goal is feasible in the sense that COMPOSER is just an alternative to text-based QENGINE programming, using the same underlying code for computations. (Additionally, similar-style programming languages such as LABVIEW is already an integral part of various experimental instrumentation control systems wherefore it can indeed also be categorized as research-enabling and this sets a precedent for realizable success). It will, however, be necessary to design and implement additional features in COMPOSER that introduces enough support and flexibility that streamlines and caters to this goal. For example, it would be beneficial to allow conditional logic nodes (If-Else, etc.), a new Code node for embedding and executing custom lines of code written by the user (direct interfacing with the underlying libraries), “locking” outputs and saving their current state in the flowscene such that lengthy computations do not need to be repeated, and so on. Further compartmentalization could also allow a user experience in close resemblance to a “programming notebook” (see e.g. Jupyter notebooks) which has become very popular in recent years due to its locally iterative and focused nature.

We may similarly envision expanding the list of educational exercises in the near-medium term future and in the longer term tailoring suites of modular, self-contained learning track trajectories. These could cover e.g. single topics or whole curricula of activities. A very ambitious and lofty goal would lie in being able to offer quantum education trajectories relevant for growing the workforce capacity related to quantum technologies and tendencies. One possible line of reasoning in this regard is as follows.

Currently, educational trajectories in quantum still largely reside in a fairly few and narrow traditional academic paths — i.e. physics and physics-adjacents — of extensive length and only a subset of these produce graduate and postgraduate candidates with the necessary skill set (or even the right “language”) to feasibly enter the quantum industry directly. One way to generate workforce capacity is thus to scale up the intake of students within these select educations, but it seems reasonable to speculate that this strategy cannot be the full solution for at least a few related reasons. First, only certain fractions of the total syllabus within these are related to quantum and are often intermittently and independently distributed without an overarching focus. Second, it is plausible that not everyone needs the full “from farm to fork” quantum generalist knowledge to contribute in solving real-world problems. This is not new thinking as many areas of chemistry, biology, nano science, etc., in particular include quantum on a sufficient “need-to-know” basis to an extent that enables meaningful

learning and research. Similar reasoning could apply to e.g. quantum technological aspirations and it might actually be beneficial to promote interdisciplinarity and diversify perspectives with complementing domain knowledge. Third, outreach and educational efforts outside strictly academic circles could be an accelerant for quantum adoption into the existing workforce and companies. While it is far beyond this thesis and personal competency to scope out and attack the full extent of these broad and complex issues, it is reasonable to at least aspire to be a part of the solution through pedagogical and visual software such as QUANTUM COMPOSER and strong accompanying didactical efforts that can convey the most central quantum ideas as frictionlessly as possible to broad population segments.

There is clearly still quite a ways to go before robustly reaching any of these obviously grand goals. QUANTUM COMPOSER is still in its relative infancy and only with continued developmental cycles, manpower, and funding will these have a chance of becoming a reality. The current version 1.0 is, viewed from a user experience perspective for example, not without its quirks, flaws, stability issues, and limitations. There are a number of yet unimplemented features that one has come to expect from high-quality software e.g. basic undo/redo functionality and similar. Future near-medium-long term releases should attempt to mend these common frustrations and allow a user experience more aligned with modern software standards. More control over visuals, plotting options, and general UX/UI should also be prioritized. Still, despite these lamentations and the constraints of limited resource allocation, QUANTUM COMPOSER's current state and achievements is not without merit and I believe there is a huge potential for it and similar software in the future.

MANUSCRIPT [6] – QUANTUM COMPOSER: A PROGRAMMABLE QUANTUM VISUALIZATION AND SIMULATION TOOL FOR EDUCATION AND RESEARCH

*Accepted for publication in American Journal of Physics.
29 pages.*

AUTHOR CONTRIBUTIONS

In the initial development phase for the first release of QUANTUM COMPOSER, I created the proof-of-concept prototype (and incidentally, during my student programming days, also the spiritual ancestor QSimTool). The ensuing software design and planning was a joint effort between mainly myself, Jens Jakob Sørensen, Till Heinzl, and Birk Skyum. The primary programmers were initially myself, Till Heinzl, and Birk Skyum. Later, Florian Korsakissok took over the programming responsibilities.

In the deployment phase, I, Shaeema Zaman, Carrie Weidner, and others, used the application in a wide range of scenarios. These include e.g. the quantum mechanics courses held by Marcel Mudrich and Jacob Sherson.

In the writing phase, I, Shaeema Zaman, and Carrie Weidner contributed equally towards first authorship. Jacob Sherson provided scientific feedback and corrections.

OUTLOOK

THE vibrant field of quantum optimal control currently finds use in many research relevant contexts and it is reasonable to believe that this trend will continue. Here we sketch out a few possible research directions in line with the topics presented throughout this thesis. Some of these were intermittently introduced to some extent, usually at the end of the chapters. Their relative feasibility and potential value may vary, yet this is paid no particular mind here.

10.1 TECHNIQUES

Chapters 4–5 showed that exact analytical derivatives can be efficiently calculated — subexponentially, gradients are approximately linear in the dynamical propagation time — for both linear and nonlinear Schrödinger equations of motion. This required simultaneous inter-linking of several analytical and numerical considerations. A natural question is then if similar developments are possible for other equations of motions.

One such example could be non-unitary dynamics [271, 419], open systems described by master equations where the density matrix is subject to dissipation, with control in the Hamiltonian as usual or in the dissipative couplings. Certain exact discretize-then-optimize gradients are well-known, but are based on exponentially expensive recursive commutators [271] and similar.

Another example could be MCDTHB dynamics [192] (and its cousins), an efficient ansatz for approximating many-body systems in terms of variationally chosen optimal single-particle modes and relative occupation numbers. As for the nonlinear Schrödinger equations, it seems that only the optimize-then-discretize is currently followed [298] and this means that the resulting derivatives are possibly approximate.

On an algorithmic level, the KROTOV methods described in Sec. 4.6 attain monotonicity by assuming continuous time as it is, to my knowledge, always derived using optimize-then-discretize. Following a discretize-then-optimize approach instead may enable monotonicity for any size of δt with a difference in update rule equivalent to Secs. 4.3–4.4.

A separate line of technique research concerns employing exact Hessians for optimizations. In general discussions we have often referred to *derivative*-based optimization since this is the most appro-

priate designation to encompass both the gradient and Hessian. Yet, presently, nearly all problems are solved with the [BFGS](#) approximations cf. [Sec. 4.1](#) which is then accurately described plainly as *gradient*-based optimization. This is quite simply because the Hessian has seldomly been worth the additional computational effort and/or prospective implementational hassles. However, although we did not delve deep into the details, it was also clear from [Ch. 4](#) that the new proposed methodologies also significantly reduces the computational effort for attaining exact Hessians, in fact much more so than for the exact gradients in relative terms. This means that employing true Newton search directions may be numerically feasible and, coupled with e.g. another work from our group [\[7\]](#) indicating potential gains even outside these new methods, it may turn out to be a relevant consideration in the future.

10.2 CONTROL PROBLEMS

The optimal control formulations in [Ch. 3](#) are generalistic and the derivative-based methods of [Chs. 4–5](#) pertain to systems of arbitrary spatial dimensionality (and Hilbert space size), so long as they follow Schrödinger-like dynamics. Yet, the control problems and simulations treated in [Chs. 5–9](#) are based on (quasi)-one-dimensional system models cf. [Secs. 2.4–2.5](#).

There are several worthwhile gains in including more spatial dimensions such as reducing/eliminating model approximation errors and enabling the study of inherently two-or-three-dimensional processes. As far as exact-gradient optimal control of such much more expansive and complex Hilbert spaces is concerned, the main computational effort still lies almost exclusively in solving the system dynamics when using the introduced methodologies. This means that, now, the feasible scope of attackable problems is now primarily reliant on how efficiently a given quantum system can be numerically treated (as opposed to being limited by approximation-deficiencies related to the control algorithms). This was one of the central points illustrated in the genuine many-body problem of [Ch. 7](#) represented by [MPS](#) and using an appropriately tailored dynamical time evolution algorithm. There, some of the possible candidates for realistically extending e.g. many-body studies, both in terms of size and dimensionality, were also highlighted (increasingly advanced tensor networks and machine-learning-type neural networks). Thus, for the future, if one or more of these endeavors (or others) prove sufficiently effective for such appropriate state description, dynamical evolution, etc., then this would open up an innumerable number of opportunities for associated control problems.

Along other lines, and in relation to our own experimental setup, great progress is being made towards enabling a variety of interesting control studies. This is due in large parts to the combination of optical lattices, digital mirror devices, and single-atom detection capabilities. The implementable models then range from one, two, to meanfield- and genuinely many-body systems — in fact all those mentioned in Sec. 2.5 and more, e.g. spintronics [420] which can be mapped onto lattice models [421–423] in the sense of quantum simulation.

One possible first exploration lies in controlling just a single particle with the versatile digital mirror device. The device can paint nearly arbitrary potentials, but is also somewhat limited in its update rate which means that the control must be e.g. “stroboscopically” parametrized (comparatively long intervals of constant control) in theoretical treatments. One could also envision joint studies of the recently proposed research problems listed in Ch. 5 related to meanfield BEC and basically also any of the other problems and related questions theoretically explored throughout the thesis.

I had personally hoped to participate in theoretical and experimental optimal control efforts during my studies but that exciting collaboration will instead be left for the next generation of students.

10.3 INTERFACES

The use of alternative interfaces for solving research problems is to my knowledge currently very limited. Our attempts to map out certain possibilities, related both to crowdsourcing as well as didactic and research-assisting aspects, have yielded initially encouraging prospects in this regard.

One of the concrete directions mentioned in Sec. 8.5 for the crowdsourcing-type interfaces lies in treating, e.g., control problems of higher complexity cf. Sec. 8.1. Methodological extrapolation remains speculative yet until comprehensively studied.

For the more didactical and research-assisting focused interfaces, one may dreamily envision the possibility transforming these into educational aids for preparing a future quantum-technology-ready workforce and standalone research-enabling instruments, respectively.

BIBLIOGRAPHY

- [1] J. J. Sørensen, J. H. M. Jensen, T. Heinzl, and J. F. Sherson. “QEngine: A C++ library for quantum optimal control of ultra-cold atoms.” In: *Computer Physics Communications* 243 (2019), pp. 135–150.
- [2] J. H. M. Jensen, J. J. Sørensen, K. Mølmer, and J. F. Sherson. “Time-optimal control of collisional $\sqrt{\text{SWAP}}$ gates in ultracold atomic systems.” In: *Physical Review A* 100 (2019), p. 052314.
- [3] J. H. M. Jensen, F. S. Møller, J. J. Sørensen, and J. F. Sherson. “Approximate Dynamics Lead to More Optimal Control: Efficient Calculation of Exact Derivatives.” In: *arXiv:2005.09943* (2020).
- [4] J. H. M. Jensen, F. S. Møller, J. J. Sørensen, and J. F. Sherson. “Achieving fast high-fidelity control of many-body dynamics.” In: *arXiv:2008.06076* (2020).
- [5] J. H. M. Jensen et al. “Crowdsourcing human common sense for quantum control.” In: *Physical Review Research* 3 (2021), p. 013057.
- [6] J. H. M. Jensen et al. “Quantum Composer: A programmable quantum visualization and simulation tool for education and research.” In: *arXiv:2006.07263* (2020).
- [7] M. Dalgaard, F. Motzoi, J. H. M. Jensen, and J. Sherson. “Hessian-based optimization of constrained quantum control.” In: *Physical Review A* 102 (2020), p. 042612.
- [8] C. A. Weidner, S. Z. Ahmed, J. H. M. Jensen, J. Sherson, and H. Lewandowski. “Investigating student use of a flexible tool for simulating and visualizing quantum mechanics.” In: *Physics Education Research Conference 2020*. 2020, pp. 563–568.
- [9] E. Schrödinger. “Are there quantum jumps? Part II.” In: *The British Journal for the Philosophy of science* 3.11 (1952), pp. 233–242.
- [10] S. Hawking. *The Dreams That Stuff Is Made Of: the most astounding papers of quantum physics – and how they shook the scientific world*. Philadelphia, PA: Running Press, 2011.
- [11] R. Omnes. *The interpretation of quantum mechanics*. Vol. 102. Princeton, NJ, USA: Princeton University Press, 2018.
- [12] M. Polymenis. “Faraday on the fiscal benefits of science.” In: *Nature* 468.7324 (2010), pp. 634–634.

- [13] J. P. Dowling and G. J. Milburn. "Quantum technology: the second quantum revolution." In: *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 361.1809 (2003), pp. 1655–1674.
- [14] T. H. Maiman. "Stimulated optical radiation in ruby." In: *Nature* 187.4736 (1960), pp. 493–494.
- [15] T. N. Theis and H.-S. P. Wong. "The end of moore's law: A new beginning for information technology." In: *Computing in Science & Engineering* 19.2 (2017), pp. 41–50.
- [16] S. Lloyd. "Universal quantum simulators." In: *Science* (1996), pp. 1073–1078.
- [17] R. P. Feynman. "There's plenty of room at the bottom." In: *California Institute of Technology, Engineering and Science magazine* (1960).
- [18] D. Barredo, V. Lienhard, S. De Leseleuc, T. Lahaye, and A. Browaeys. "Synthetic three-dimensional atomic structures assembled atom by atom." In: *Nature* 561.7721 (2018), pp. 79–82.
- [19] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell. "Observation of Bose-Einstein condensation in a dilute atomic vapor." In: *Science* 269.5221 (1995), pp. 198–201.
- [20] C. C. Bradley, C. Sackett, J. Tollett, and R. G. Hulet. "Evidence of Bose-Einstein condensation in an atomic gas with attractive interactions." In: *Physical Review Letters* 75.9 (1995), p. 1687.
- [21] K. B. Davis et al. "Bose-Einstein condensation in a gas of sodium atoms." In: *Physical Review Letters* 75.22 (1995), p. 3969.
- [22] O Astafiev et al. "Resonance fluorescence of a single artificial atom." In: *Science* 327.5967 (2010), pp. 840–843.
- [23] J. You and F. Nori. "Atomic physics and quantum optics using superconducting circuits." In: *Nature* 474.7353 (2011), pp. 589–597.
- [24] I. Buluta, S. Ashhab, and F. Nori. "Natural and artificial atoms for quantum computation." In: *Reports on Progress in Physics* 74.10 (2011), p. 104401.
- [25] R. Hanson and D. D. Awschalom. "Coherent manipulation of single spins in semiconductors." In: *Nature* 453.7198 (2008), pp. 1043–1049.
- [26] C. H. Bennett et al. "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels." In: *Physical Review Letters* 70.13 (1993), p. 1895.
- [27] D. Bouwmeester et al. "Experimental quantum teleportation." In: *Nature* 390.6660 (1997), pp. 575–579.

- [28] J. F. Sherson et al. "Quantum teleportation between light and matter." In: *Nature* 443.7111 (2006), pp. 557–560.
- [29] R. Valivarthi et al. "Teleportation Systems Toward a Quantum Internet." In: *PRX Quantum* 1.2 (2020), p. 020317.
- [30] R. P. Feynman. "Simulating physics with computers." In: *Int. J. Theor. Phys* 21.6/7 (1982).
- [31] A. Steane. "Quantum computing." In: *Reports on Progress in Physics* 61.2 (1998), p. 117.
- [32] C. H. Bennett and D. P. DiVincenzo. "Quantum information and computation." In: *Nature* 404.6775 (2000), pp. 247–255.
- [33] D. P. DiVincenzo. "The physical implementation of quantum computation." In: *Fortschritte der Physik: Progress of Physics* 48.9-11 (2000), pp. 771–783.
- [34] H. Häffner, C. F. Roos, and R. Blatt. "Quantum computing with trapped ions." In: *Physics reports* 469.4 (2008), pp. 155–203.
- [35] J. Preskill. "Quantum Computing in the NISQ era and beyond." In: *Quantum* 2 (2018), p. 79.
- [36] A. A. Houck, H. E. Türeci, and J. Koch. "On-chip quantum simulation with superconducting circuits." In: *Nature Physics* 8.4 (2012), pp. 292–299.
- [37] R. Blatt and C. F. Roos. "Quantum simulations with trapped ions." In: *Nature Physics* 8.4 (2012), pp. 277–284.
- [38] I. M. Georgescu, S. Ashhab, and F. Nori. "Quantum simulation." In: *Reviews of Modern Physics* 86.1 (2014), p. 153.
- [39] C. Gross and I. Bloch. "Quantum simulations with ultracold atoms in optical lattices." In: *Science* 357.6355 (2017), pp. 995–1001.
- [40] A. Acín et al. "The quantum technologies roadmap: a European community view." In: *New Journal of Physics* 20.8 (2018), p. 080201.
- [41] R. Thew, T. Jennewein, and M. Sasaki. "Focus on quantum science and technology initiatives around the world." In: *Quantum Science and Technology* 4 (2019).
- [42] E. R. MacQuarrie, C. Simon, S. Simmons, and E. Maine. "The emerging commercial landscape of quantum computing." In: *Nature Reviews Physics* 2.11 (2020), pp. 596–598.
- [43] M. Mohseni et al. "Commercialize quantum technologies in five years." In: *Nature News* 543.7644 (2017), p. 171.
- [44] M. Riedel, M. Kovacs, P. Zoller, J. Mlynek, and T. Calarco. "Europe's quantum flagship initiative." In: *Quantum Science and Technology* 4.2 (2019), p. 020501.

- [45] M. G. Raymer and C. Monroe. "The US national quantum initiative." In: *Quantum Science and Technology* 4.2 (2019), p. 020504.
- [46] Q. Zhang, F. Xu, L. Li, N.-L. Liu, and J.-W. Pan. "Quantum information research in China." In: *Quantum Science and Technology* 4.4 (2019), p. 040503.
- [47] D. d'Alessandro. *Introduction to quantum control and dynamics*. Boca Raton, FL: CRC press, 2007.
- [48] W. S. Widnall. "The minimum-time thrust-vector control law in the Apollo lunar-module autopilot." In: *IFAC Proceedings Volumes* 3.1 (1970), pp. 136–153.
- [49] D. J. Tannor and S. A. Rice. "Control of selectivity of chemical reaction via control of wave packet evolution." In: *The Journal of chemical physics* 83.10 (1985), pp. 5013–5018.
- [50] C. Brif, R. Chakrabarti, and H. Rabitz. "Control of quantum phenomena: past, present and future." In: *New Journal of Physics* 12.7 (2010), p. 075008.
- [51] M. Shapiro and P. Brumer. *Quantum control of molecular processes*. John Wiley & Sons, 2012.
- [52] S. J. Glaser et al. "Training Schrödinger's cat: quantum optimal control." In: *The European Physical Journal D* 69.12 (2015), p. 279.
- [53] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm. "Simple pulses for elimination of leakage in weakly nonlinear qubits." In: *Physical Review Letters* 103.11 (2009), p. 110501.
- [54] D. J. Egger and F. K. Wilhelm. "Optimized controlled-z gates for two superconducting qubits coupled through a resonator." In: *Superconductor Science and Technology* 27.1 (2013), p. 014001.
- [55] M. H. Goerz, F. Motzoi, K. B. Whaley, and C. P. Koch. "Charting the circuit QED design landscape using optimal control theory." In: *npj Quantum Information* 3.1 (2017), pp. 1–10.
- [56] S. Montangero. *Introduction to Tensor Network Methods: Numerical simulations of low-dimensional many-body quantum systems*. Springer International Publishing, 2018. ISBN: 9783030014094.
- [57] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson. "Global optimization of quantum dynamics with AlphaZero deep exploration." In: *npj Quantum Information* 6.1 (2020), p. 6.
- [58] C. T. Kehlet et al. "Improving solid-state NMR dipolar recoupling by optimal control." In: *Journal of the American Chemical Society* 126.33 (2004), pp. 10202–10203.
- [59] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser. "Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms." In: *Journal of Magnetic Resonance* 172.2 (2005), pp. 296–305.

- [60] N. C. Nielsen, C. Kehlet, S. J. Glaser, and N. Khaneja. "Optimal control methods in NMR spectroscopy." In: *eMagRes* (2007).
- [61] W. Kallies and S. J. Glaser. "Cooperative broadband spin echoes through optimal control." In: *Journal of Magnetic Resonance* 286 (2018), pp. 115–137.
- [62] J. J. Sørensen, J. S. Nyemann, F. Motzoi, J. Sherson, and T. Vosegaard. "Optimization of pulses with low bandwidth for improved excitation of multiple-quantum coherences in NMR of quadrupolar nuclei." In: *The Journal of Chemical Physics* 152.5 (2020), p. 054104.
- [63] J. Scheuer et al. "Precise qubit control beyond the rotating wave approximation." In: *New Journal of Physics* 16.9 (2014), p. 093022.
- [64] F. Dolde et al. "High-fidelity spin entanglement using optimal control." In: *Nature Communications* 5.1 (2014), pp. 1–9.
- [65] G. Waldherr et al. "Quantum error correction in a solid-state hybrid spin register." In: *Nature* 506.7487 (2014), pp. 204–207.
- [66] Y. Chou, S.-Y. Huang, and H.-S. Goan. "Optimal control of fast and high-fidelity quantum gates with electron and nuclear spins of a nitrogen-vacancy center in diamond." In: *Physical Review A* 91.5 (2015), p. 052315.
- [67] V. Nebendahl, H. Häffner, and C. Roos. "Optimal control of entangling operations for trapped-ion quantum computing." In: *Physical Review A* 79.1 (2009), p. 012312.
- [68] C. P. Koch, J. P. Palao, R. Kosloff, and F. Masnou-Seeuws. "Stabilization of ultracold molecules using optimal control theory." In: *Physical Review A* 70.1 (2004), p. 013402.
- [69] C. P. Koch, E. Luc-Koenig, and F. Masnou-Seeuws. "Making ultracold molecules in a two-color pump-dump photoassociation scheme using chirped pulses." In: *Physical Review A* 73.3 (2006), p. 033408.
- [70] E. F. De Lima, T.-S. Ho, and H. Rabitz. "Optimal laser control of molecular photoassociation along with vibrational stabilization." In: *Chemical Physics Letters* 501.4-6 (2011), pp. 267–272.
- [71] K. M. Tibbetts, X. Xing, and H. Rabitz. "Optimal control of molecular fragmentation with homologous families of photonic reagents and chemical substrates." In: *Physical Chemistry Chemical Physics* 15.41 (2013), pp. 18012–18022.
- [72] G. De Chiara et al. "Optimal control of atom transport for quantum gates in optical lattices." In: *Physical Review A* 77.5 (2008), p. 052333.

- [73] P. Doria, T. Calarco, and S. Montangero. "Optimal control technique for many-body quantum dynamics." In: *Physical Review Letters* 106.19 (2011), p. 190501.
- [74] S. van Frank et al. "Optimal control of complex atomic quantum systems." In: *Scientific reports* 6 (2016), p. 34187.
- [75] G. Jäger and U. Hohenester. "Optimal quantum control of Bose-Einstein condensates in magnetic microtraps: Consideration of filter effects." In: *Physical Review A* 88.3 (2013), p. 035601.
- [76] J. Cui, R. van Bijnen, T. Pohl, S. Montangero, and T. Calarco. "Optimal control of Rydberg lattice gases." In: *Quantum Science and Technology* 2.3 (2017), p. 035006.
- [77] S. Patsch et al. "Fast and accurate circularization of a Rydberg atom." In: *Physical Review A* 97.5 (2018), p. 053418.
- [78] A. Larrouy et al. "Fast Navigation in a Large Hilbert Space Using Quantum Optimal Control." In: *Physical Review X* 10.2 (2020), p. 021058.
- [79] K. Bongs et al. "Taking atom interferometric quantum sensors from the laboratory to real-world applications." In: *Nature Reviews Physics* (2019), pp. 1–9.
- [80] M. G. Bason et al. "Measurement-enhanced determination of BEC phase transitions." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 51.17 (2018), p. 175301.
- [81] R. Heck et al. "Remote optimization of an ultracold atoms experiment by experts and citizen scientists." In: *Proceedings of the National Academy of Sciences* 115.48 (2018), E11231–E11237. ISSN: 0027-8424.
- [82] O. Eliasson et al. "Spatially-selective in situ magnetometry of ultracold atomic clouds." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 52.7 (2019), p. 075003.
- [83] O. Eliasson et al. "Spatial tomography of individual atoms in a quantum gas microscope." In: *Physical Review A* 102.5 (2020), p. 053311.
- [84] O. Eliasson. "Magnetometry and microscopy of cold atom clouds." PhD thesis. Aarhus University, 2020.
- [85] J. S. Laustsen. "Spatial tomography of single atoms in a quantum gas microscope." PhD thesis. Aarhus University, 2020.
- [86] I. Bloch, J. Dalibard, and W. Zwerger. "Many-body physics with ultracold gases." In: *Reviews of Modern Physics* 80.3 (2008), p. 885.
- [87] O. Morsch and M. Oberthaler. "Dynamics of Bose-Einstein condensates in optical lattices." In: *Reviews of Modern Physics* 78.1 (2006), p. 179.

- [88] S. Giorgini, L. P. Pitaevskii, and S. Stringari. "Theory of ultra-cold atomic Fermi gases." In: *Reviews of Modern Physics* 80.4 (2008), p. 1215.
- [89] M. Lewenstein, A. Sanpera, and V. Ahufinger. *Ultracold Atoms in Optical Lattices: Simulating quantum many-body systems*. Oxford, UK: Oxford University Press, 2012.
- [90] W. Ketterle, D. S. Durfee, and D. Stamper-Kurn. "Making, probing and understanding Bose-Einstein condensates." In: *arXiv preprint cond-mat/9904034* (1999).
- [91] R. Grimm, M. Weidemüller, and Y. B. Ovchinnikov. "Optical Dipole Traps for Neutral Atoms." In: ed. by B. Bederson and H. Walther. Vol. 42. *Advances In Atomic, Molecular, and Optical Physics*. Academic Press, 2000, pp. 95–170.
- [92] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller. "Cold bosonic atoms in optical lattices." In: *Physical Review Letters* 81.15 (1998), p. 3108.
- [93] D. Stamper-Kurn et al. "Optical confinement of a Bose-Einstein condensate." In: *Physical Review Letters* 80.10 (1998), p. 2027.
- [94] J. F. Sherson et al. "Single-atom-resolved fluorescence imaging of an atomic Mott insulator." In: *Nature* 467.7311 (2010), p. 68.
- [95] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th. New York, NY, USA: Cambridge University Press, 2011. ISBN: 1107002176, 9781107002173.
- [96] A. Montanaro. "Quantum algorithms: an overview." In: *npj Quantum Information* 2.1 (2016), pp. 1–8.
- [97] S. Jordan. *The quantum algorithm zoo*. Available at <https://quantumalgorithmzoo.org>.
- [98] S. Boixo et al. "Characterizing quantum supremacy in near-term devices." In: *Nature Physics* 14.6 (2018), pp. 595–600.
- [99] I. Georgescu. "The DiVincenzo criteria 20 years on." In: *Nature Reviews Physics* (2020), pp. 1–1.
- [100] D. A. Lidar, I. L. Chuang, and K. B. Whaley. "Decoherence-free subspaces for quantum computation." In: *Physical Review Letters* 81.12 (1998), p. 2594.
- [101] T. Schulte-Herbrüggen, A. Spörl, N. Khaneja, and S. Glaser. "Optimal control for generating quantum gates in open dissipative systems." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 44.15 (2011), p. 154013.
- [102] D. A. Lidar and T. A. Brun. *Quantum error correction*. New York, NY, USA: Cambridge university press, 2013.
- [103] E. T. Campbell, B. M. Terhal, and C. Vuillot. "Roads towards fault-tolerant universal quantum computation." In: *Nature* 549.7671 (2017), pp. 172–179.

- [104] T. D. Ladd et al. "Quantum computers." In: *Nature* 464.7285 (2010), pp. 45–53.
- [105] V. Cerletti, W. Coish, O. Gywat, and D. Loss. "Recipes for spin-based quantum computing." In: *Nanotechnology* 16.4 (2005), R27.
- [106] H.-J. Briegel, T. Calarco, D. Jaksch, J. I. Cirac, and P. Zoller. "Quantum computing with neutral atoms." In: *Journal of modern optics* 47.2-3 (2000), pp. 415–451.
- [107] S. Chu. "Cold atoms and quantum control." In: *Nature* 416.6877 (2002), p. 206.
- [108] C. Weitenberg et al. "Single-spin addressing in an atomic Mott insulator." In: *Nature* 471.7338 (2011), p. 319.
- [109] A. Kaufman et al. "Two-particle quantum interference in tunnel-coupled optical tweezers." In: *Science* 345.6194 (2014), pp. 306–309.
- [110] Y. Wang, X. Zhang, T. A. Corcovilos, A. Kumar, and D. S. Weiss. "Coherent addressing of individual neutral atoms in a 3D optical lattice." In: *Physical Review Letters* 115.4 (2015), p. 043003.
- [111] H. Kim et al. "In situ single-atom array synthesis using dynamic holographic optical tweezers." In: *Nature communications* 7 (2016), p. 13317.
- [112] M. Endres et al. "Atom-by-atom assembly of defect-free one-dimensional cold atom arrays." In: *Science* 354.6315 (2016), pp. 1024–1027.
- [113] D. Barredo, S. De Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys. "An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays." In: *Science* 354.6315 (2016), pp. 1021–1023.
- [114] W. Lee, H. Kim, and J. Ahn. "Three-dimensional rearrangement of single atoms using actively controlled optical microtraps." In: *Optics express* 24.9 (2016), pp. 9816–9825.
- [115] A. Kumar, T.-Y. Wu, F. Giraldo, and D. S. Weiss. "Sorting ultracold atoms in a three-dimensional optical lattice in a realization of Maxwell's demon." In: *Nature* 561.7721 (2018), p. 83.
- [116] S. Sashkin, J. Wilson, B. Grinkemeyer, and J. D. Thompson. "Narrow-line cooling and imaging of ytterbium atoms in an optical tweezer array." In: *Physical Review Letters* 122.14 (2019), p. 143002.
- [117] M. Norcia, A. Young, and A. Kaufman. "Microscopic control and detection of ultracold strontium in optical-tweezer arrays." In: *Physical Review X* 8.4 (2018), p. 041054.

- [118] A. Cooper et al. "Alkaline-earth atoms in optical tweezers." In: *Physical Review X* 8.4 (2018), p. 041055.
- [119] L. Henriët et al. "Quantum computing with neutral atoms." In: *Quantum* 4 (2020), p. 327.
- [120] K. Eckert et al. "Quantum computing in optical microtraps based on the motional states of neutral atoms." In: *Physical Review A* 66.4 (2002), p. 042317.
- [121] J. Mompart, K. Eckert, W. Ertmer, G. Birkl, and M. Lewenstein. "Quantum computing with spatially delocalized qubits." In: *Physical Review Letters* 90.14 (2003), p. 147901.
- [122] E. Charron, E. Tiesinga, F. Mies, and C. Williams. "Optimizing a phase gate using quantum interference." In: *Physical Review Letters* 88.7 (2002), p. 077901.
- [123] E. Charron, M. Cirone, A. Negretti, J. Schmiedmayer, and T. Calarco. "Theoretical analysis of a realistic atom-chip quantum gate." In: *Physical Review A* 74.1 (2006), p. 012308.
- [124] G. K. Brennen, C. M. Caves, P. S. Jessen, and I. H. Deutsch. "Quantum logic gates in optical lattices." In: *Physical Review Letters* 82.5 (1999), p. 1060.
- [125] D. Jaksch et al. "Fast Quantum Gates for Neutral Atoms." In: *Physical Review Letters* 85.10 (Sept. 2000), pp. 2208–2211.
- [126] M. Lukin et al. "Dipole blockade and quantum information processing in mesoscopic atomic ensembles." In: *Physical Review Letters* 87.3 (2001), p. 037901.
- [127] A. J. Daley, M. M. Boyd, J. Ye, and P. Zoller. "Quantum computing with alkaline-earth-metal atoms." In: *Physical Review Letters* 101.17 (2008), p. 170504.
- [128] A. Negretti, P. Treutlein, and T. Calarco. "Quantum computing implementations with neutral particles." In: *Quantum information processing* 10.6 (2011), p. 721.
- [129] C. Weitenberg, S. Kuhr, K. Mølmer, and J. F. Sherson. "Quantum computation architecture using optical tweezers." In: *Physical Review A* 84.3 (2011), p. 032322.
- [130] P.-I. Schneider and A. Saenz. "Quantum computation with ultracold atoms in a driven optical lattice." In: *Physical Review A* 85.5 (2012), p. 050304.
- [131] N. B. Jørgensen, M. G. Bason, and J. F. Sherson. "One- and two-qubit quantum gates using superimposed optical-lattice potentials." In: *Phys. Rev. A* 89 (3 2014), p. 032306.
- [132] G. Pagano, F. Scazza, and M. Foss-Feig. "Fast and Scalable Quantum Information Processing with Two-Electron Atoms in Optical Tweezer Arrays." In: *Advanced Quantum Technologies* 2.3-4 (2019), p. 1800067.

- [133] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch. "Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms." In: *Nature* 415.6867 (2002), pp. 39–44.
- [134] S. Lloyd. "Almost any quantum logic gate is universal." In: *Physical Review Letters* 75.2 (1995), p. 346.
- [135] A. Barenco et al. "Elementary gates for quantum computation." In: *Physical Review A* 52.5 (1995), p. 3457.
- [136] I. H. Deutsch, G. K. Brennen, and P. S. Jessen. "Quantum Computing with Neutral Atoms in an Optical Lattice." In: *Fortschritte der Physik* 48.9-11 (Sept. 2000), pp. 925–943. ISSN: 1521-3978.
- [137] T Xia et al. "Randomized benchmarking of single-qubit gates in a 2D array of neutral-atom qubits." In: *Physical Review Letters* 114.10 (2015), p. 100503.
- [138] Y. Wang, A. Kumar, T.-Y. Wu, and D. S. Weiss. "Single-qubit gates based on targeted phase shifts in a 3D neutral atom array." In: *Science* 352.6293 (2016), pp. 1562–1565.
- [139] O. Mandel et al. "Controlled collisions for multi-particle entanglement of optically trapped atoms." In: *Nature* 425.6961 (2003), p. 937.
- [140] M. Anderlini et al. "Controlled exchange interaction between pairs of neutral atoms in an optical lattice." In: *Nature* 448.7152 (2007), pp. 452–456.
- [141] T. Wilk et al. "Entanglement of two individual neutral atoms using Rydberg blockade." In: *Physical Review Letters* 104.1 (2010), p. 010502.
- [142] X. Zhang, L. Isenhower, A. Gill, T. Walker, and M Saffman. "Deterministic entanglement of two neutral atoms via Rydberg blockade." In: *Physical Review A* 82.3 (2010), p. 030306.
- [143] L Isenhower et al. "Demonstration of a neutral atom controlled-NOT quantum gate." In: *Physical Review Letters* 104.1 (2010), p. 010503.
- [144] K. Maller et al. "Rydberg-blockade controlled-not gate and entanglement in a two-dimensional array of neutral-atom qubits." In: *Physical Review A* 92.2 (2015), p. 022336.
- [145] A. Kaufman et al. "Entangling two transportable neutral atoms via local spin exchange." In: *Nature* 527.7577 (2015), p. 208.
- [146] Y.-Y. Jau, A. Hankin, T. Keating, I. Deutsch, and G. Biedermann. "Entangling atomic spins with a Rydberg-dressed spin-flip blockade." In: *Nature Physics* 12.1 (2016), p. 71.

- [147] H. Levine et al. "High-fidelity control and entanglement of rydberg-atom qubits." In: *Physical Review Letters* 121.12 (2018), p. 123603.
- [148] T. Schaetz, C. R. Monroe, and T. Esslinger. "Focus on quantum simulation." In: *New Journal of Physics* 15.8 (2013), p. 085009.
- [149] S. Sachdev. *Quantum Phase Transitions*. Cambridge: Cambridge University Press, 2011. ISBN: 9780521514682.
- [150] J. I. Cirac and P. Zoller. "Goals and opportunities in quantum simulation." In: *Nature Physics* 8.4 (2012), pp. 264–266.
- [151] F. Tacchino, A. Chiesa, S. Carretta, and D. Gerace. "Quantum Computers as Universal Quantum Simulators: State-of-the-Art and Perspectives." In: *Advanced Quantum Technologies* 3.3 (2020), p. 1900052.
- [152] M. Kjaergaard et al. "Superconducting qubits: Current state of play." In: *Annual Review of Condensed Matter Physics* 11 (2020), pp. 369–395.
- [153] J. I. Cirac. "Quantum computing and simulation: Where we stand and what awaits us." In: *Nanophotonics* 10.1 (2020), pp. 453–456.
- [154] N. K. Lysne, K. W. Kuper, P. M. Poggi, I. H. Deutsch, and P. S. Jessen. "Small, Highly Accurate Quantum Processor for Intermediate-Depth Quantum Simulations." In: *Physical Review Letters* 124.23 (2020), p. 230501.
- [155] A. de Touzalin et al. *Quantum manifesto: a new era of technology*. 2016.
- [156] L. V. Hau, S. E. Harris, Z. Dutton, and C. H. Behroozi. "Light speed reduction to 17 metres per second in an ultracold atomic gas." In: *Nature* 397.6720 (1999), pp. 594–598.
- [157] J. Klaers, J. Schmitt, F. Vewinger, and M. Weitz. "Bose–Einstein condensation of photons in an optical microcavity." In: *Nature* 468.7323 (2010), pp. 545–548.
- [158] C. J. Pethick and H. Smith. *Bose-Einstein condensation in dilute gases*. New York, NY, USA: Cambridge university press, 2008.
- [159] L. Pitaevskii and S. Stringari. *Bose-Einstein condensation and superfluidity*. Vol. 164. Oxford University Press, 2016.
- [160] T Schumm et al. "Matter-wave interferometry in a double well on an atom chip." In: *Nature Physics* 1.1 (2005), pp. 57–62.
- [161] U. Hohenester, P. K. Rekdal, A. Borzì, and J. Schmiedmayer. "Optimal quantum control of Bose-Einstein condensates in magnetic microtraps." In: *Physical Review A* 75.2 (2007), p. 023602.

- [162] G. Jäger, D. M. Reich, M. H. Goerz, C. P. Koch, and U. Hohenester. "Optimal quantum control of Bose-Einstein condensates in magnetic microtraps: Comparison of gradient-ascent-pulse-engineering and Krotov optimization schemes." In: *Physical Review A* 90.3 (2014), p. 033628.
- [163] J. J. Sørensen, M. Aranburu, T. Heinzl, and J. Sherson. "Approaching the Quantum Speed Limit with Global-Local Optimization." In: *arXiv:1802.07521* (2018).
- [164] R. Bücke et al. "Twin-atom beams." In: *Nature Physics* 7.8 (2011), pp. 608–611.
- [165] J. J. Sørensen, M. O. Aranburu, T. Heinzl, and J. Sherson. "Quantum optimal control in a chopped basis: Applications in control of Bose-Einstein condensates." In: *Physical Review A* 98.2 (2018), p. 022119.
- [166] F Borselli et al. "Two-particle Interference in Double Twin-atom-beams." In: *arXiv:2009.13438v1* (2020).
- [167] J.-F. Mennemann, T. Langen, L. Exl, and N. J. Mauser. "Optimal control of the self-bound dipolar droplet formation process." In: *Computer Physics Communications* 244 (2019), pp. 205–216.
- [168] S. Chu, J. E. Bjorkholm, A. Ashkin, and A. Cable. "Experimental Observation of Optically Trapped Atoms." In: *Physical Review Letters* 57.3 (1986), pp. 314–317.
- [169] N. Schlosser, G. Reymond, I. Protsenko, and P. Grangier. "Subpoissonian loading of single atoms in a microscopic dipole trap." In: *Nature* 411.6841 (2001), pp. 1024–1027.
- [170] P. Zupancic et al. "Ultra-precise holographic beam shaping for microscopic quantum control." In: *Optics express* 24.13 (2016), pp. 13881–13893.
- [171] M. Anderlini, J. Sebby-Strabley, J. Kruse, J. V. Porto, and W. D. Phillips. "Controlled atom dynamics in a double-well optical lattice." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 39.10 (2006), S199.
- [172] I Lesanovsky et al. "Adiabatic radio-frequency potentials for the coherent manipulation of matter waves." In: *Physical Review A* 73.3 (2006), p. 033619.
- [173] J. Fortágh and C. Zimmermann. "Magnetic microtraps for ultracold atoms." In: *Reviews of Modern Physics* 79.1 (2007), p. 235.
- [174] T. L. Harte et al. "Ultracold atoms in multiple radio-frequency dressed adiabatic potentials." In: *Physical Review A* 97.1 (2018), p. 013616.

- [175] B. M. Garraway and H. Perrin. "Recent developments in trapping and manipulation of atoms with adiabatic potentials." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 49.17 (2016), p. 172001.
- [176] R. Folman, P. Krüger, J. Schmiedmayer, J. Denschlag, and C. Henkel. "Microscopic Atom Optics: From Wires to an Atom Chip." In: ed. by B. Bederson and H. Walther. Vol. 48. *Advances In Atomic, Molecular, and Optical Physics*. Academic Press, 2002, pp. 263–356.
- [177] S. Hofferberth, I. Lesanovsky, B. Fischer, T. Schumm, and J. Schmiedmayer. "Non-equilibrium coherence dynamics in one-dimensional Bose gases." In: *Nature* 449.7160 (2007), pp. 324–327.
- [178] F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari. "Theory of Bose-Einstein condensation in trapped gases." In: *Reviews of Modern Physics* 71.3 (1999), p. 463.
- [179] A. Altland and B. D. Simons. *Condensed matter field theory*. New York, NY, USA: Cambridge university press, 2010.
- [180] H. Bruus and K. Flensberg. "Many-body quantum theory in condensed matter physics." In: Oxford: Oxford University Press, 2004.
- [181] J. J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Cambridge: Cambridge University Press, 2017. ISBN: 9781108422413.
- [182] J. Weiner, V. S. Bagnato, S. Zilio, and P. S. Julienne. "Experiments and theory in cold and ultracold collisions." In: *Reviews of Modern Physics* 71.1 (1999), p. 1.
- [183] N. R. Thomas, N. Kjærgaard, P. S. Julienne, and A. C. Wilson. "Imaging of s and d partial-wave interference in quantum scattering of identical bosonic atoms." In: *Physical Review Letters* 93.17 (2004), p. 173201.
- [184] C.-K. Hong, Z.-Y. Ou, and L. Mandel. "Measurement of subpicosecond time intervals between two photons by interference." In: *Physical Review Letters* 59.18 (1987), p. 2044.
- [185] R. Liu, B. Odom, Y. Yamamoto, and S. Tarucha. "Quantum interference in electron collision." In: *Nature* 391.6664 (1998), pp. 263–265.
- [186] P. S. Julienne, F. Mies, E. Tiesinga, and C. J. Williams. "Collisional stability of double Bose condensates." In: *Physical Review Letters* 78.10 (1997), p. 1880.
- [187] A. I. Streltsov, O. E. Alon, and L. S. Cederbaum. "Formation and dynamics of many-boson fragmented states in one-dimensional attractive ultracold gases." In: *Physical Review Letters* 100.13 (2008), p. 130401.

- [188] D Baillie, R. Wilson, R. Bisset, and P. Blakie. "Self-bound dipolar droplet: A localized matter wave in free space." In: *Physical Review A* 94.2 (2016), p. 021602.
- [189] A. U. Lode. "Multiconfigurational time-dependent Hartree method for bosons with internal degrees of freedom: Theory and composite fragmentation of multicomponent Bose-Einstein condensates." In: *Physical Review A* 93.6 (2016), p. 063601.
- [190] J. Nguyen et al. "Parametric excitation of a Bose-Einstein condensate: From Faraday waves to granulation." In: *Physical Review X* 9.1 (2019), p. 011052.
- [191] A. I. Streltsov, O. E. Alon, and L. S. Cederbaum. "Role of excited states in the splitting of a trapped interacting Bose-Einstein condensate by a time-dependent barrier." In: *Physical Review Letters* 99.3 (2007), p. 030402.
- [192] O. E. Alon, A. I. Streltsov, and L. S. Cederbaum. "Multiconfigurational time-dependent Hartree method for bosons: Many-body dynamics of bosonic systems." In: *Physical Review A* 77.3 (2008), p. 033613.
- [193] R. Lin et al. "MCTDH-X: The multiconfigurational time-dependent Hartree method for indistinguishable particles software." In: *Quantum Science and Technology* 5.2 (2020), p. 024004.
- [194] A. U. Lode, S. Dutta, and C. L  v  que. "Dynamics of Ultra-cold Bosons in Artificial Gauge Fields: Angular Momentum, Fragmentation, and the Variance of Entropy." In: *arXiv preprint arXiv:2012.09870* (2020).
- [195] I. Brouzos et al. "Quantum speed limit and optimal control of many-boson dynamics." In: *Physical Review A* 92.6 (2015), p. 062110.
- [196] A. U. Lode, C. L  v  que, L. B. Madsen, A. I. Streltsov, and O. E. Alon. "Colloquium: Multiconfigurational time-dependent Hartree approaches for indistinguishable particles." In: *Reviews of Modern Physics* 92.1 (2020), p. 011001.
- [197] J.-F. Mennemann, D. Matthes, R.-M. Weish  upl, and T. Langen. "Optimal control of Bose-Einstein condensates in three dimensions." In: *New Journal of Physics* 17.11 (2015), p. 113027.
- [198] F. L. Claude Cohen-Tannoudji Bernard Diu. *Quantum Mechanics, Volume 1*. New York, NY, USA: Wiley, 1991.
- [199] M. Olshanii. "Atomic scattering in the presence of an external confinement and a gas of impenetrable bosons." In: *Physical Review Letters* 81.5 (1998), p. 938.

- [200] L. D. Carr, M. A. Leung, and W. P. Reinhardt. "Dynamics of the Bose-Einstein condensate: quasi-one-dimension and beyond." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 33.19 (2000), p. 3983.
- [201] T. Busch, B.-G. Englert, K. Rzażewski, and M. Wilkens. "Two Cold Atoms in a Harmonic Trap." In: *Foundations of Physics* 28.4 (1998), pp. 549–559. ISSN: 1572-9516.
- [202] P. Kościk and T. Sowiński. "Exactly solvable model of two trapped quantum particles interacting via finite-range soft-core interactions." In: *Scientific Reports* 8 (2018).
- [203] R. A. Doganov, S. Klaiman, O. E. Alon, A. I. Streltsov, and L. S. Cederbaum. "Two trapped particles interacting by a finite-range two-body potential in two spatial dimensions." In: *Phys. Rev. A* 87 (3 2013), p. 033631.
- [204] J. Freericks and H Monien. "Phase diagram of the Bose-Hubbard model." In: *EPL (Europhysics Letters)* 26.7 (1994), p. 545.
- [205] M. Tsang. "Quantum transition-edge detectors." In: *Physical Review A* 88.2 (2013), p. 021801.
- [206] P. Zanardi, P. Giorda, and M. Cozzini. "Information-theoretic differential geometry of quantum phase transitions." In: *Physical Review Letters* 99.10 (2007), p. 100603.
- [207] P. Barmettler and C. Kollath. "Quantum gases in optical lattices." In: *Quantum Gas Experiments: Exploring Many-Body States*. World Scientific, 2015, pp. 33–61.
- [208] M. Weidemüller and C. Zimmermann. *Interactions in ultracold gases: from atoms to molecules*. John Wiley & Sons, 2011.
- [209] O. S. Sørensen. "Exact treatment of interacting bosons in rotating systems and lattices." PhD thesis. Aarhus University, Department of Physics and Astronomy, 2012.
- [210] W. Xu, M. Olshanii, and M. Rigol. "Multiband effects and the Bose-Hubbard model in one-dimensional lattices." In: *Physical Review A* 94.3 (2016), p. 031601.
- [211] O. Goldreich. *P, NP, and NP-Completeness: The basics of computational complexity*. New York, NY, USA: Cambridge University Press, 2010.
- [212] D. Basilewitsch, C. P. Koch, and D. M. Reich. "Quantum Optimal Control for Mixed State Squeezing in Cavity Optomechanics." In: *Advanced Quantum Technologies* 2.3-4 (2019), p. 1800110.
- [213] S. G. Schirmer and P. de Fouquieres. "Efficient algorithms for optimal control of quantum dynamics: the Krotov method unencumbered." In: *New Journal of Physics* 13.7 (2011), p. 073029.
- [214] J. Nocedal and S. J. Wright. *Numerical optimization 2nd*. Springer Science & Business Media, New York, 2006.

- [215] D. Liberzon. *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press, 2011.
- [216] A. Borzì, G. Ciaramella, and M. Sprengel. *Formulation and Numerical Solution of Quantum Control Problems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2017, p. 397. ISBN: 978-1-61197-483-6.
- [217] R. Chakrabarti and H. Rabitz. “Quantum control landscapes.” In: *International Reviews in Physical Chemistry* 26.4 (2007), pp. 671–735.
- [218] M. Larocca, P. M. Poggi, and D. A. Wisniacki. “Quantum control landscape for a two-level system near the quantum speed limit.” In: *Journal of Physics A: Mathematical and Theoretical* 51.38 (2018), p. 385305.
- [219] A. Nanduri, A. Donovan, T.-S. Ho, and H. Rabitz. “Exploring quantum control landscape structure.” In: *Phys. Rev. A* 88 (3 2013), p. 033425.
- [220] K. W. Moore, R. Chakrabarti, G. Riviello, and H. Rabitz. “Search complexity and resource scaling for the quantum optimal control of unitary transformations.” In: *Phys. Rev. A* 83 (1 2011), p. 012326.
- [221] M. Bukov et al. “Broken symmetry in a two-qubit quantum control landscape.” In: *Phys. Rev. A* 97 (5 2018), p. 052114.
- [222] G. C. Hegerfeldt. “Driving at the quantum speed limit: optimal control of a two-level system.” In: *Physical Review Letters* 111.26 (2013), p. 260501.
- [223] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm. “Tunable, flexible, and efficient optimization of control pulses for practical qubits.” In: *Physical Review Letters* 120.15 (2018), p. 150401.
- [224] M. H. Goerz et al. “Krotov: A Python implementation of Krotov’s method for quantum optimal control.” In: *SciPost physics* 7 (2019).
- [225] L. B. Rall and G. F. Corliss. “An introduction to automatic differentiation.” In: *Computational Differentiation: Techniques, Applications, and Tools* 89 (1996).
- [226] N. Leung, M. Abdelhafez, J. Koch, and D. Schuster. “Speedup for quantum optimal control from automatic differentiation based on graphics processing units.” In: *Physical Review A* 95.4 (2017), p. 042318.
- [227] M. Črepinšek, S.-H. Liu, and M. Mernik. “Exploration and exploitation in evolutionary algorithms: A survey.” In: *ACM computing surveys (CSUR)* 45.3 (2013), pp. 1–33.

- [228] H. A. Rabitz, M. M. Hsieh, and C. M. Rosenthal. "Quantum optimally controlled transition landscapes." In: *Science* 303.5666 (2004), pp. 1998–2001.
- [229] K. W. Moore and H. Rabitz. "Exploring constrained quantum control landscapes." In: *The Journal of chemical physics* 137.13 (2012), p. 134113.
- [230] H. Rabitz, T.-S. Ho, R. Long, R. Wu, and C. Brif. "Comment on: are there traps in quantum control landscapes?" In: *Physical Review Letters* 108.19 (2012), p. 198901.
- [231] A. N. Pechen and D. J. Tannor. "Are there traps in quantum control landscapes?" In: *Physical Review Letters* 106.12 (2011), p. 120402.
- [232] P. De Fouquieres and S. G. Schirmer. "A closer look at quantum control landscapes and their implication for control optimization." In: *Infinite dimensional analysis, quantum probability and related topics* 16.03 (2013), p. 1350021.
- [233] Z. Ugray et al. "Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization." In: *INFORMS Journal on Computing* 19.3 (2007), pp. 328–340.
- [234] S. Sivanandam and S. Deepa. "Genetic algorithms." In: *Introduction to genetic algorithms*. Berlin, Heidelberg: Springer, 2008, pp. 15–37.
- [235] P. J. Van Laarhoven and E. H. Aarts. "Simulated annealing." In: *Simulated annealing: Theory and applications*. Springer, 1987, pp. 7–15.
- [236] Q. Bai. "Analysis of particle swarm optimization algorithm." In: *Computer and information science* 3.1 (2010), p. 180.
- [237] D. M. Olsson and L. S. Nelson. "The Nelder-Mead simplex procedure for function minimization." In: *Technometrics* 17.1 (1975), pp. 45–51.
- [238] E. Zahedinejad, S. Schirmer, and B. C. Sanders. "Evolutionary algorithms for hard quantum control." In: *Physical Review A* 90.3 (2014), p. 032310.
- [239] P. J. Liebermann and F. K. Wilhelm. "Optimal qubit control using single-flux quantum pulses." In: *Physical Review Applied* 6.2 (2016), p. 024022.
- [240] M. G. Bason et al. "High-fidelity quantum driving." In: *Nature Physics* 8.2 (2012), pp. 147–152.
- [241] T. Caneva, T. Calarco, and S. Montangero. "Chopped random-basis quantum optimization." In: *Physical Review A* 84.2 (2011), p. 022326.

- [242] N. Rach, M. M. Müller, T. Calarco, and S. Montangero. "Dressing the chopped-random-basis optimization: A bandwidth-limited access to the trap-free landscape." In: *Physical Review A* 92.6 (2015), p. 062343.
- [243] Y.-X. Du et al. "Experimental realization of stimulated Raman shortcut-to-adiabatic passage with cold atoms." In: *Nature communications* 7.1 (2016), pp. 1–7.
- [244] R. Fisher. "Optimal control of multi-level quantum systems." PhD thesis. Technische Universität München, 2010.
- [245] F. A. Potra and S. J. Wright. "Interior-point methods." In: *Journal of Computational and Applied Mathematics* 124.1-2 (2000), pp. 281–302.
- [246] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization." In: *ACM Transactions on Mathematical Software (TOMS)* 23.4 (1997), pp. 550–560.
- [247] S. G. Schirmer, H. Fu, and A. I. Solomon. "Complete controllability of quantum systems." In: *Physical Review A* 63.6 (2001), p. 063410.
- [248] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon. "Optimizing variational quantum algorithms using Pontryagin's minimum principle." In: *Physical Review X* 7.2 (2017), p. 021027.
- [249] A. Chefles. "Quantum state discrimination." In: *Contemporary Physics* 41.6 (2000), pp. 401–424.
- [250] S. Deffner and S. Campbell. "Quantum speed limits: from Heisenbergs uncertainty principle to optimal quantum control." In: *Journal of Physics A: Mathematical and Theoretical* 50.45 (2017), p. 453001.
- [251] L. Mandelstam and I. Tamm. "The uncertainty relation between energy and time in non-relativistic quantum mechanics." In: *J. Phys.* 9 (1945), p. 249.
- [252] N. Margolus and L. B. Levitin. "The maximum speed of dynamical evolution." In: *Physica D: Nonlinear Phenomena* 120.1-2 (1998), pp. 188–195.
- [253] L. B. Levitin and T. Toffoli. "Fundamental limit on the rate of quantum dynamics: the unified bound is tight." In: *Physical Review Letters* 103.16 (2009), p. 160502.
- [254] P. Král, I. Thanopoulos, and M. Shapiro. "Colloquium: Coherently controlled adiabatic passage." In: *Reviews of Modern Physics* 79.1 (2007), p. 53.

- [255] K. Bergmann et al. "Roadmap on STIRAP applications." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 52.20 (2019), p. 202001.
- [256] H. L. Mortensen, J. J. W. Sørensen, K. Mølmer, and J. F. Sherson. "Fast state transfer in a Λ -system: a shortcut-to-adiabaticity approach to robust and resource optimized control." In: *New Journal of Physics* 20.2 (2018), p. 025009.
- [257] D. J. Egger and F. K. Wilhelm. "Adaptive hybrid optimal quantum control for imprecisely characterized systems." In: *Physical Review Letters* 112.24 (2014), p. 240503.
- [258] N. Wittler et al. "An integrated tool-set for Control, Calibration and Characterization of quantum devices applied to superconducting qubits." In: *arXiv:2009.09866* (2020).
- [259] D. H. Wolpert and W. G. Macready. "No free lunch theorems for optimization." In: *IEEE transactions on evolutionary computation* 1.1 (1997), pp. 67–82.
- [260] M. Avriel. *Nonlinear programming: analysis and methods*. Courier Corporation, 2003.
- [261] D. P. Bertsekas, W. Hager, and O. Mangasarian. *Nonlinear programming*. Athena Scientific Belmont, MA, 1998.
- [262] M. Frank, P. Wolfe, et al. "An algorithm for quadratic programming." In: *Naval research logistics quarterly* 3.1-2 (1956), pp. 95–110.
- [263] D. Goodwin and I. Kuprov. "Modified Newton-Raphson GRAPE methods for optimal control of spin systems." In: *The Journal of chemical physics* 144.20 (2016), p. 204107.
- [264] G. Von Winckel and A. Borzì. "Computational techniques for a quantum control problem with H_1 -cost." In: *Inverse Problems* 24.3 (2008), p. 034007.
- [265] A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming." In: *Mathematical Programming* 106.1 (2006), pp. 25–57.
- [266] J. J. Moré and D. J. Thuente. "Line Search Algorithms with Guaranteed Sufficient Decrease." In: *ACM Trans. Math. Softw.* 20.3 (Sept. 1994), pp. 286–307. ISSN: 0098-3500.
- [267] G. Fasano, G. Liuzzi, S. Lucidi, and F. Rinaldi. "A linesearch-based derivative-free approach for nonsmooth constrained optimization." In: *SIAM journal on optimization* 24.3 (2014), pp. 959–992.

- [268] D.-H. Li and M. Fukushima. "A derivative-free line search and global convergence of Broyden-like method for nonlinear equations." In: *Optimization methods and software* 13.3 (2000), pp. 181–201.
- [269] H. Zhang and W. W. Hager. "A nonmonotone line search technique and its application to unconstrained optimization." In: *SIAM journal on Optimization* 14.4 (2004), pp. 1043–1056.
- [270] W. W. Hager and H. Zhang. "A new conjugate gradient method with guaranteed descent and an efficient line search." In: *SIAM Journal on optimization* 16.1 (2005), pp. 170–192.
- [271] P De Fouquieres, S. Schirmer, S. Glaser, and I. Kuprov. "Second order gradient ascent pulse engineering." In: *Journal of Magnetic Resonance* 212.2 (2011), pp. 412–417.
- [272] J. J. Sørensen. "Quantum Control of Ultracold Quantum Systems." PhD thesis. Aarhus University, 2018.
- [273] J Werschnik and E. Gross. "Quantum optimal control theory." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 40.18 (2007), R175.
- [274] D. M. Reich, M. Ndong, and C. P. Koch. "Monotonically convergent optimization in quantum control using Krotov's method." In: *The Journal of chemical physics* 136.10 (2012), p. 104103.
- [275] U. Schollwöck. "The density-matrix renormalization group in the age of matrix product states." In: *Annals of Physics* 326.1 (2011), pp. 96–192.
- [276] F. F. Floether, P. De Fouquieres, and S. G. Schirmer. "Robust quantum gates for open systems via optimal control: Markovian versus non-Markovian dynamics." In: *New Journal of Physics* 14.7 (2012), p. 073023.
- [277] D. Goodwin and I. Kuprov. "Auxiliary matrix formalism for interaction representation transformations, optimal control, and spin relaxation theories." In: *The Journal of chemical physics* 143.8 (2015), p. 084113.
- [278] R. Beerwerth and H. Bauke. "Krylov subspace methods for the Dirac equation." In: *Computer Physics Communications* 188 (2015), pp. 189–197.
- [279] H. Fehske et al. "Numerical approaches to time evolution of complex quantum systems." In: *Physics Letters A* 373.25 (2009), pp. 2182–2188.
- [280] D. Lauvergnat, S. Blasco, X. Chapuisat, and A. Nauts. "A simple and efficient evolution operator for time-dependent Hamiltonians: the Taylor expansion." In: *The Journal of chemical physics* 126.20 (2007), p. 204103.

- [281] A. Alvermann, H Fehske, and P. Littlewood. "Numerical time propagation of quantum systems in radiation fields." In: *New Journal of Physics* 14.10 (2012), p. 105008.
- [282] S. Savitz and G. Refael. "Stable unitary integrators for the numerical implementation of continuous unitary transformations." In: *Physical Review B* 96.11 (2017), p. 115129.
- [283] S Machnes et al. "Comparing, optimizing, and benchmarking quantum-control algorithms in a unifying programming framework." In: *Physical Review A* 84.2 (2011), p. 022305.
- [284] H. Hogben, M Krzystyniak, G. Charnock, P. Hore, and I. Kuprov. "Spinach—a software library for simulation of spin dynamics in large spin systems." In: *Journal of Magnetic Resonance* 208.2 (2011), pp. 179–194.
- [285] J. R. Johansson, P. D. Nation, and F. Nori. "QuTiP: An open-source Python framework for the dynamics of open quantum systems." In: *Computer Physics Communications* 183.8 (2012), pp. 1760–1772.
- [286] A. P. Peirce, M. A. Dahleh, and H. Rabitz. "Optimal control of quantum-mechanical systems: Existence, numerical approximation, and applications." In: *Physical Review A* 37.12 (1988), p. 4950.
- [287] A. Konnov and V. F. Krotov. "On global methods for the successive improvement of control processes." In: *Avtomatika i Telemekhanika* 10 (1999), pp. 77–88.
- [288] R. Eitan, M. Mundt, and D. J. Tannor. "Optimal control with accelerated convergence: Combining the Krotov and quasi-Newton methods." In: *Physical Review A* 83.5 (2011), p. 053426.
- [289] S. E. Sklarz and D. J. Tannor. "Loading a Bose-Einstein condensate onto an optical lattice: An application of optimal control theory to the nonlinear Schrödinger equation." In: *Physical Review A* 66.5 (2002), p. 053619.
- [290] B. Riaz, C. Shuang, and S. Qamar. "Optimal control methods for quantum gate preparation: a comparative study." In: *Quantum Information Processing* 18.4 (2019), pp. 1–26.
- [291] S.-H. Wu, M. Amezcua, and H. Wang. "Adiabatic population transfer of dressed spin states with quantum optimal control." In: *Physical Review A* 99.6 (2019), p. 063812.
- [292] F. K. Wilhelm, S. Kirchhoff, S. Machnes, N. Wittler, and D. Sugny. "An introduction into optimal control for quantum technologies." In: *arXiv:2003.10132* (2020).
- [293] U. Boscain, M. Sigalotti, and D. Sugny. "Introduction to the Foundations of Quantum Optimal Control." In: *arXiv preprint arXiv:2010.09368* (2020).

- [294] R. Mukherjee, H. Xie, and F. Mintert. "Bayesian Optimal Control of Greenberger-Horne-Zeilinger States in Rydberg Lattices." In: *Physical Review Letters* 125.20 (2020), p. 203603.
- [295] Ž. Marojević, E. Göklü, and C. Lämmerzahl. "Energy eigenfunctions of the 1D Gross-Pitaevskii equation." In: *Computer Physics Communications* 184.8 (2013), pp. 1920–1930.
- [296] M. Bonneau, W. J. Munro, K. Nemoto, and J. Schmiedmayer. "Characterizing twin-particle entanglement in double-well potentials." In: *Physical Review A* 98.3 (2018), p. 033608.
- [297] R. Bücker et al. "Vibrational state inversion of a Bose-Einstein condensate: optimal control and state tomography." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 46.10 (2013), p. 104012.
- [298] U. Hohenester. "OCTBEC - A Matlab toolbox for optimal quantum control of Bose-Einstein condensates." In: *Computer Physics Communications* 185.1 (2014), pp. 194–216.
- [299] M. Mundt and D. J. Tannor. "Optimal control of interacting particles: a multi-configuration time-dependent Hartree-Fock approach." In: *New Journal of Physics* 11.10 (2009), p. 105038.
- [300] M. Saffman. *The next step in making arrays of single atoms*. 2019.
- [301] A. J. Daley. "Quantum computing and quantum simulation with group-II atoms." In: *Quantum Information Processing* 10.6 (2011), p. 865.
- [302] M. Theis et al. "Tuning the scattering length with an optically induced Feshbach resonance." In: *Physical Review Letters* 93.12 (2004), p. 123001.
- [303] E. Urban et al. "Observation of Rydberg blockade between two atoms." In: *Nature Physics* 5.2 (2009), pp. 110–114.
- [304] M. Saffman, T. G. Walker, and K. Mølmer. "Quantum information with Rydberg atoms." In: *Reviews of Modern Physics* 82.3 (2010), p. 2313.
- [305] H. Bernien et al. "Probing many-body dynamics on a 51-atom quantum simulator." In: *Nature* 551.7682 (2017), pp. 579–584.
- [306] A. Omran et al. "Generation and manipulation of Schrödinger cat states in Rydberg atom arrays." In: *Science* 365.6453 (2019), pp. 570–574.
- [307] W. Li. "A boost to Rydberg quantum computing." In: *Nature Physics* (2020), pp. 1–2.
- [308] S. Ebadi et al. "Quantum Phases of Matter on a 256-Atom Programmable Quantum Simulator." In: *arXiv:2012.12281* (2020).
- [309] P. Scholl et al. "Programmable quantum simulation of 2D anti-ferromagnets with hundreds of Rydberg atoms." In: *arXiv:2012.12268* (2020).

- [310] Y. Zeng et al. "Entangling two individual atoms of different isotopes via Rydberg blockade." In: *Physical Review Letters* 119.16 (2017), p. 160502.
- [311] X.-F. Shi. "Deutsch, toffoli, and cnot gates via rydberg blockade of neutral atoms." In: *Physical Review Applied* 9.5 (2018), p. 051001.
- [312] P. Weinberg and M. Bukov. "QuSpin: a Python package for dynamics and exact diagonalisation of quantum many body systems part I: spin chains." In: (2017).
- [313] S. Iskakov and M. Danilov. "Exact diagonalization library for quantum electron models." In: *Computer Physics Communications* 225 (2018), pp. 128–139.
- [314] A. Weiße and H. Fehske. "Exact diagonalization techniques." In: *Computational many-particle physics*. Springer, 2008, pp. 529–544.
- [315] J. Zhang and R. Dong. "Exact diagonalization: the Bose–Hubbard model as an example." In: *European Journal of Physics* 31.3 (2010), p. 591.
- [316] D. Raventós, T. Graß, M. Lewenstein, and B. Juliá-Díaz. "Cold bosons in optical lattices: a tutorial for exact diagonalization." In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 50.11 (2017), p. 113001.
- [317] I Foster et al. "High Performance Eigenvalue Solver for Hubbard Model: Tuning Strategies for LOBPCG Method on CUDA GPU." In: *Parallel Computing: Technology Trends* 36 (2020), p. 105.
- [318] D. Poulin, A. Qarry, R. Somma, and F. Verstraete. "Quantum simulation of time-dependent Hamiltonians and the convenient illusion of Hilbert space." In: *Physical Review Letters* 106.17 (2011), p. 170501.
- [319] R. Orús. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states." In: *Annals of Physics* 349 (2014), pp. 117–158.
- [320] J. Eisert and T. J. Osborne. "General entanglement scaling laws from time evolution." In: *Physical Review Letters* 97.15 (2006), p. 150404.
- [321] J. Eisert, M. Cramer, and M. B. Plenio. "Colloquium: Area laws for the entanglement entropy." In: *Reviews of Modern Physics* 82.1 (2010), p. 277.
- [322] P. Calabrese and J. Cardy. "Entanglement entropy and quantum field theory." In: *Journal of Statistical Mechanics: Theory and Experiment* 2004.06 (2004), Po6002.

- [323] P. Calabrese and J. Cardy. "Entanglement entropy and quantum field theory: a non-technical introduction." In: *International Journal of Quantum Information* 4.03 (2006), pp. 429–438.
- [324] G. Vidal. "Efficient simulation of one-dimensional quantum many-body systems." In: *Physical Review Letters* 93.4 (2004), p. 040502.
- [325] G. Vidal. "Efficient classical simulation of slightly entangled quantum computations." In: *Physical Review Letters* 91.14 (2003), p. 147902.
- [326] S. Lloyd and S. Montangero. "Information theoretical analysis of quantum optimal control." In: *Physical Review Letters* 113.1 (2014), p. 010502.
- [327] J. Eisert, M. Friesdorf, and C. Gogolin. "Quantum many-body systems out of equilibrium." In: *Nature Physics* 11.2 (2015), pp. 124–130.
- [328] F. Verstraete, V. Murg, and J. I. Cirac. "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems." In: *Advances in Physics* 57.2 (2008), pp. 143–224.
- [329] U. Schollwöck. "The density-matrix renormalization group." In: *Reviews of Modern Physics* 77.1 (2005), p. 259.
- [330] G. Evenbly and G. Vidal. "Tensor network states and geometry." In: *Journal of Statistical Physics* 145.4 (2011), pp. 891–918.
- [331] M. C. Bañuls et al. "Simulating lattice gauge theories within quantum technologies." In: *The European physical journal D* 74.8 (2020), pp. 1–42.
- [332] T. Felser, P. Silvi, M. Collura, and S. Montangero. "Two-dimensional quantum-link lattice quantum electrodynamics at finite density." In: *Physical Review X* 10.4 (2020), p. 041040.
- [333] G. Magnifico, T. Felser, P. Silvi, and S. Montangero. "Lattice Quantum Electrodynamics in $(3+1)$ -dimensions at finite density with Tensor Networks." In: *arXiv:2011.10658* (2020).
- [334] In: *ITensor Library (version 2)* <http://itensor.org> ().
- [335] J. J. Sakurai and E. D. Commins. *Modern quantum mechanics, revised edition*. 1995.
- [336] C. Hubig. "Symmetry-Protected Tensor Networks." PhD thesis. LMU Munich, 2017.
- [337] S. Paeckel et al. "Time-evolution methods for matrix-product states." In: *Annals of Physics* 411 (2019), p. 167998.
- [338] A. J. Daley, C. Kollath, U. Schollwöck, and G. Vidal. "Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces." In: *Journal of Statistical Mechanics: Theory and Experiment* 2004.04 (2004), P04005.

- [339] M. P. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher. "Boson localization and the superfluid-insulator transition." In: *Physical Review B* 40.1 (1989), p. 546.
- [340] T. D. Kühner and H. Monien. "Phases of the one-dimensional Bose-Hubbard model." In: *Physical Review B* 58.22 (1998), R14741.
- [341] T. D. Kühner, S. R. White, and H. Monien. "One-dimensional Bose-Hubbard model with nearest-neighbor interaction." In: *Physical Review B* 61.18 (2000), p. 12474.
- [342] M. Lewenstein et al. "Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond." In: *Advances in Physics* 56.2 (2007), pp. 243–379.
- [343] C. Kollath, A. M. Läuchli, and E. Altman. "Quench dynamics and nonequilibrium phase diagram of the Bose-Hubbard model." In: *Physical Review Letters* 98.18 (2007), p. 180601.
- [344] T. Caneva, T. Calarco, R. Fazio, G. E. Santoro, and S. Montangero. "Speeding up critical system dynamics through optimized evolution." In: *Physical Review A* 84.1 (2011), p. 012312.
- [345] D. Quiñones-Valles, S. Dolgov, and D. Savostyanov. "Tensor Product Approach to Quantum Control." In: *Integral Methods in Science and Engineering: Analytic Treatment and Numerical Approximations*. Ed. by C. Constanda and P. Harris. Cham: Springer International Publishing, 2019, pp. 367–379. ISBN: 978-3-030-16077-7.
- [346] B. Damski. "The simplest quantum model supporting the Kibble-Zurek mechanism of topological defect production: Landau-Zener transitions from a new perspective." In: *Physical Review Letters* 95.3 (2005), p. 035701.
- [347] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car. "Theory of quantum annealing of an Ising spin glass." In: *Science* 295.5564 (2002), pp. 2427–2430.
- [348] T. Caneva et al. "Optimal control at the quantum speed limit." In: *Physical Review Letters* 103.24 (2009), p. 240501.
- [349] T. Caneva et al. "Complexity of controlling quantum many-body dynamics." In: *Physical Review A* 89.4 (2014), p. 042322.
- [350] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac. "Classical simulation of infinite-size quantum lattice systems in two spatial dimensions." In: *Physical Review Letters* 101.25 (2008), p. 250602.
- [351] L. Carr. *Understanding quantum phase transitions*. CRC press, 2010.

- [352] J. O. Iregui, M. Troyer, and P. Corboz. "Infinite matrix product states versus infinite projected entangled-pair states on the cylinder: A comparative study." In: *Physical Review B* 96.11 (2017), p. 115113.
- [353] A Kshetrimayum, M Rizzi, J. Eisert, and R Orús. "Tensor network annealing algorithm for two-dimensional thermal states." In: *Physical Review Letters* 122.7 (2019), p. 070502.
- [354] G. Carleo and M. Troyer. "Solving the quantum many-body problem with artificial neural networks." In: *Science* 355.6325 (2017), pp. 602–606.
- [355] D.-L. Deng, X. Li, and S. D. Sarma. "Quantum entanglement in neural network states." In: *Physical Review X* 7.2 (2017), p. 021021.
- [356] J. Biamonte et al. "Quantum machine learning." In: *Nature* 549.7671 (2017), pp. 195–202.
- [357] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac. "Restricted Boltzmann machines in quantum physics." In: *Nature Physics* 15.9 (2019), pp. 887–892.
- [358] M. Dalgaard, F. Motzoi, J. J. Sorensen, and J. Sherson. "Global optimization of quantum dynamics with AlphaZero deep exploration." In: *npj Quantum Information* 6.1 (2020).
- [359] G. Carleo et al. "Machine learning and the physical sciences." In: *Reviews of Modern Physics* 91.4 (2019), p. 045002.
- [360] A Bohrdt et al. "Analyzing non-equilibrium quantum states through snapshots with artificial neural networks." In: *arXiv:2012.11586* (2020).
- [361] F. Marquardt. "Machine Learning and Quantum Devices." In: *arXiv:2101.01759* (2021).
- [362] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feed-forward networks are universal approximators." In: *Neural Networks* 2.5 (1989), pp. 359–366.
- [363] M. Collura, L. Dell'Anna, T. Felser, and S. Montangero. "On the descriptive power of Neural-Networks as constrained Tensor Networks with exponentially large bond dimension." In: *arXiv:1905.11351* (2019).
- [364] G. Torlai et al. "Neural-network quantum state tomography." In: *Nature Physics* 14.5 (2018), pp. 447–450.
- [365] J. Silvertown. "A new dawn for citizen science." In: *Trends in ecology & evolution* 24.9 (2009), pp. 467–471.
- [366] R. Bonney et al. "Citizen science: a developing tool for expanding science knowledge and scientific literacy." In: *BioScience* 59.11 (2009), pp. 977–984.

- [367] L. See et al. "Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information." In: *ISPRS International Journal of Geo-Information* 5.5 (2016), p. 55.
- [368] L. Von Ahn and L. Dabbish. "Labeling images with a computer game." In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, pp. 319–326.
- [369] C. Abellán et al. "Challenging local realism with human choices." In: *Nature* 557.7704 (2018), pp. 212–216.
- [370] C. J. Lintott et al. "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." In: *Monthly Notices of the Royal Astronomical Society* 389.3 (2008), pp. 1179–1189.
- [371] A. Kawrykow et al. "Phylo: a citizen science approach for improving multiple sequence alignment." In: *PloS one* 7.3 (2012), e31362.
- [372] J. Waldispühl, A. Szantner, R. Knight, S. Caisse, and R. Pitchford. "Leveling up citizen science." In: *Nature Biotechnology* 38.10 (2020), pp. 1124–1126.
- [373] J. S. Kim et al. "Space–time wiring specificity supports direction selectivity in the retina." In: *Nature* 509.7500 (2014), pp. 331–336.
- [374] J. Lee et al. "RNA design rules from a massive open laboratory." In: *Proceedings of the National Academy of Sciences* 111.6 (2014), pp. 2122–2127.
- [375] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl. "The protein folding problem." In: *Annu. Rev. Biophys.* 37 (2008), pp. 289–316.
- [376] S. Cooper et al. "Predicting protein structures with a multiplayer online game." In: *Nature* 466.7307 (2010), pp. 756–760.
- [377] B. Koepnick et al. "De novo protein design by citizen scientists." In: *Nature* 570.7761 (2019), pp. 390–394.
- [378] J. Howe. "The rise of crowdsourcing." In: *Wired magazine* 14.6 (2006), pp. 1–4.
- [379] P. Michelucci and J. L. Dickinson. "The power of crowds." In: *Science* 351.6268 (2016), pp. 32–33.
- [380] D. Dörner and J. Funke. "Complex problem solving: what it is and what it is not." In: *Frontiers in psychology* 8 (2017), p. 1153.
- [381] W. E. Forum. *Future of Jobs Report*. 2018.
- [382] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu. "Deep blue." In: *Artificial intelligence* 134.1-2 (2002), pp. 57–83.

- [383] J. Burmeister and J. Wiles. "The challenge of Go as a domain for AI research: a comparison between Go and chess." In: *Proceedings of Third Australian and New Zealand Conference on Intelligent Information Systems. ANZIIS-95*. IEEE. 1995, pp. 181–186.
- [384] D. Silver et al. "Mastering the game of Go with deep neural networks and tree search." In: *Nature* 529.7587 (2016), pp. 484–489.
- [385] D. Silver et al. "Mastering the game of go without human knowledge." In: *Nature* 550.7676 (2017), pp. 354–359.
- [386] J. Reason. *Human error*. Cambridge university press, 1990.
- [387] D. Heaven. "Why deep-learning AIs are so easy to fool." In: *Nature* 574.7777 (2019), p. 163.
- [388] G. Marcus. "Deep learning: A critical appraisal." In: *arXiv preprint arXiv:1801.00631* (2018).
- [389] G. Marcus. "The next decade in AI: four steps towards robust artificial intelligence." In: *arXiv:2002.06177* (2020).
- [390] G. J. Fletcher. "Psychology and Common Sense." In: *IEEE Engineering Management Review* 14.4 (1986), pp. 30–40.
- [391] R. Jia and P. Liang. "Adversarial examples for evaluating reading comprehension systems." In: *arXiv preprint arXiv:1707.07328* (2017).
- [392] D. Kahneman. *Thinking, Fast and Slow*. Macmillan, 2011.
- [393] E. Kamar. "Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence." In: *IJCAI*. 2016, pp. 4070–4073.
- [394] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister. "Hybrid intelligence." In: *Business & Information Systems Engineering* 61.5 (2019), pp. 637–643.
- [395] A Berditchevskaia and P Baeck. *Future of Minds and Machines, Nesta Report*. 2020.
- [396] E. Baltz et al. "Achievement of sustained net plasma heating in a fusion experiment with the optometrist algorithm." In: *Scientific reports* 7.1 (2017), pp. 1–7.
- [397] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir. "The future of human-in-the-loop cyber-physical systems." In: *Computer* 46.1 (2013), pp. 36–45.
- [398] L. F. Cranor. "A framework for reasoning about the human in the loop." In: (2008).
- [399] G. Li. "Human-in-the-loop data integration." In: *Proceedings of the VLDB Endowment* 10.12 (2017), pp. 2006–2017.

- [400] A. Endert et al. "The human is the loop: new directions for visual analytics." In: *Journal of intelligent information systems* 43.3 (2014), pp. 411–435.
- [401] L. Rothrock and S Narayanan. *Human-in-the-loop simulations*. Springer, 2011.
- [402] C. Herodotou, M. Aristeidou, M. Sharples, and E. Scanlon. "Designing citizen science tools for learning: lessons learnt from the iterative development of nQuire." In: *Research and Practice in Technology Enhanced Learning* 13.1 (2018), pp. 1–23.
- [403] R. Tinati et al. "Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform." In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, pp. 4069–4078.
- [404] C. Jennett and A. Cox. "Eight guidelines for designing virtual citizen science projects." In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. Vol. 2. 1. 2014.
- [405] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In: *KDD96* (1996), pp. 226–231.
- [406] F. Khatib et al. "Algorithm discovery by protein folding game players." In: *Proceedings of the National Academy of Sciences* 108.47 (2011), pp. 18949–18953.
- [407] J. Rafner et al. "crea. blender: A Neural Network-Based Image Generation Game to Assess Creativity." In: *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. 2020, pp. 340–344.
- [408] S. Cooper, A. L. Sterling, R. Kleffner, W. M. Silversmith, and J. B. Siegel. "Repurposing citizen science games as software tools for professional scientists." In: *Proceedings of the 13th International Conference on the Foundations of Digital Games*. 2018, pp. 1–6.
- [409] J. R. Hiller, I. D. Johnston, and I. Johnson. *Quantum Mechanics Simulations*. John Wiley & Sons, Inc., 1995.
- [410] D. A. Zollman, N. S. Rebello, and K. Hogg. "Quantum mechanics for everyone: Hands-on activities integrated with technology." In: *American Journal of Physics* 70.3 (2002), pp. 252–259.
- [411] M. Belloni and W. Christian. "Physlets for quantum mechanics." In: *Computing in Science & Engineering* 5.1 (2003), pp. 90–97.
- [412] L. T. Escalada, N. S. Rebello, and D. A. Zollman. "Student explorations of quantum effects in LEDs and luminescent devices." In: *The Physics Teacher* 42.3 (2004), pp. 173–179.

- [413] B. Thaller. *Visual quantum mechanics: selected topics with computer-generated animations of quantum-mechanical phenomena*. Springer Science & Business Media, 2007.
- [414] S. McKagan et al. "Developing and researching PhET simulations for teaching quantum mechanics." In: *American Journal of Physics* 76.4 (2008), pp. 406–417.
- [415] C. Singh. "Interactive learning tutorials on quantum mechanics." In: *American Journal of Physics* 76.4 (2008), pp. 400–405.
- [416] A. Kohnle et al. "A new multimedia resource for teaching quantum mechanics concepts." In: *American Journal of Physics* 80.2 (2012), pp. 148–153.
- [417] A. Kohnle, C. Baily, A. Campbell, N. Korolkova, and M. J. Paetkau. "Enhancing student learning of two-level quantum systems with interactive simulations." In: *American Journal of Physics* 83.6 (2015), pp. 560–566.
- [418] A. Kohnle, A. Jackson, and M. Paetkau. "The difference between a probability and a probability density." In: *The Physics Teacher* 57.3 (2019), pp. 190–192.
- [419] M. H. Goerz, D. M. Reich, and C. P. Koch. "Optimal control theory for a unitary operation under dissipative evolution." In: *New Journal of Physics* 16.5 (2014), p. 055012.
- [420] S. G. Schirmer, E. A. Jonckheere, and F. C. Langbein. "Design of feedback control laws for information transfer in spintronics networks." In: *IEEE Transactions on Automatic Control* 63.8 (2017), pp. 2523–2536.
- [421] L.-M. Duan, E. Demler, and M. D. Lukin. "Controlling Spin Exchange Interactions of Ultracold Atoms in Optical Lattices." In: *Phys. Rev. Lett.* 91 (9 2003), p. 090402.
- [422] S. Trotzky et al. "Time-Resolved Observation and Control of Superexchange Interactions with Ultracold Atoms in Optical Lattices." In: *Science* 319.5861 (2008), pp. 295–299. ISSN: 0036-8075.
- [423] J. Simon et al. "Quantum simulation of antiferromagnetic spin chains in an optical lattice." In: *Nature* 472.7343 (2011), pp. 307–312.

COLOPHON

All figures were made with `MATLAB` or `Blender3D`.

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst’s seminal book on typography “*The Elements of Typographic Style*”. `classicthesis` is available for both $\text{\LaTeX}$ and `LyX`:

<https://bitbucket.org/amiede/classicthesis/>

Final Version as of January 31, 2021 (`classicthesis` v4.6).